The Evaluation of Models Designed to Represent the Surface Energy Balance of Urban Environments

Best, Martin John

Awarding institution:
King's College London

The Evaluation of Models Designed to Represent the Surface Energy Balance of Urban Environments

Martin John Best

Department of Geography
King's College London

Submitted to the University of London for the degree of Doctor of Philosophy

July 2015

Abstract

Although urban models have been individually evaluated in numerous studies, it is not possible to objectively compare the performance between models in general as the level of optimisation during each model study is often not published. Also observational datasets are often of short duration and hence the schemes have not necessarily been tested over a seasonal cycle. The first urban model comparison project was designed to address both of these issues.

Results from the comparison show that multiple reflections of shortwave radiation within street canyons, the reduction in the amount of visible sky from within the canyon (which impacts on the net long-wave radiation), the contrast in surface temperatures between building roofs and street canyons, and evaporation from vegetation are the dominant physical processes that govern the turbulent sensible and latent heat flux exchange between an urban surface and the atmospheric boundary layer. Hence models that use an appropriate bulk albedo, represent building roof surfaces separately from street canyons and include a representation of vegetation demonstrate the most skill. So models need to include these characteristics, but further complexity does not add to the skill of the model at the main study site used for the comparison, for minimizing errors in the turbulent sensible and latent heat fluxes. Furthermore, these models require parameter information on the bulk albedo, the height to width ratio of the building and the roof to impervious surface fraction, and plan area vegetation fraction.

Urban models that contributed to the comparison are able to capture the seasonal cycle and the observed trends in the fluxes with respect to atmospheric forcing. However, initial conditions of soil moisture are critical because of the importance of evaporation from vegetation.

Including the anthropogenic heat flux in the urban surface energy balance can give improved simulations, but the signal is small in the results for the comparison because of the small magnitude of the heat flux itself. Larger impacts are anticipated for urban environments with larger anthropogenic heat fluxes. Anthropogenic water sources from irrigation and street cleaning also have a critical impact on the surface energy balance, and it can be assumed that vegetation is never water stressed unless there is an irrigation ban enforced.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to thank Andy Coutts, Jason Beringer and Nigel Tapper for providing the Melbourne data that were used in the model comparison, along with everyone who participated by providing model simulations. Without their support there would have been no results to base this study upon. I would also like to thank Andreas Christen and Leena Järvi for allowing use of their various datasets that I used for the study of varying vegetation fractions.

I would also like to thank the Met Office for providing me with the opportunity to undertake this PhD and for providing the funding. My thanks go to both of my supervisors, Prof. Sue Grimmond and Dr. Andreas Baas, for their help and advice, but in particular to Sue for allowing me to use the many urban datasets that she has worked on and also for her patience during times when it has been difficult to make progress. Finally I would like to thank Karen, my wife, for her support throughout the PhD which has been challenging along side a full time role at the Met Office.

Chapter 1: Introduction

1.1 Background

The majority of the World's population now lives in urban areas, and this is set to increase to almost 70% by 2050 (World Urbanization Prospects, 2009). This means that there is a growing requirement to provide accurate weather information for the urban environment to inform and protect these inhabitants. The dispersion of pollutants within the city, that are generated by the human activities such as transportation and heating, need to be accurately predicted to reduce the impact on the health of the urban dwellers. In addition, the impacts of climate change on these urban environments are not well known, yet it is important to understand these impacts in order to adapt to any changes in the climatic conditions by future design and use of technology. Recent flooding events in cities (e.g., Neal et al., 2009, Coulthard and Frostick, 2010) have highlighted the need to understand not only the effects of heat and momentum within these environments, but also the impact of the water cycle.

From a global perspective, the build environment makes up only a small percentage of the land surface. Traditionally with the numerical models used for weather forecasting and climate prediction, cities are a sub-grid-scale feature of the land cover and are hence not resolved. This means that the land surface schemes used within these models have neglected the impact of both the urban areas on the atmospheric evolution, and the change in meteorological conditions within the city compared to the surrounding rural areas (e.g., Viterbo and Beljaars, 1995, Chen et al., 1996, Polcher et al., 1996). However, modern advances in computing power have enabled the resolution of these models to increase to the extent that cities are now being resolved in the high resolution limited area models, and even within the global models for some of the larger urban environments around the World (e.g., Best, 2006).

The latest generation of land surface models have adopted the tile or mosaic approach to represent the heterogeneity of the land surface within a numerical grid-box (e.g., Avissar and Pielke, 1989, Koster and Suarez, 1992, Essery et al., 2003). This means that these schemes calculate the exchange of heat, moisture and momentum from a number of different surface types and blend these fluxes at a suitable height using a linear weighting that depends upon the fractional coverage for each land cover that constitutes the grid-box. As such, the models are capable of representing sub-grid-scale features such as cities. This has recently led to increasing interest from the weather community in how to develop their models to include urban surfaces.

Operational weather forecasts require not only a large amount of computing to complete the simulations, but also need to be delivered in a timely fashion. This means that any scientific development within the schemes that constitute the weather forecasting model are required to be computationally efficient, whilst retaining the required level of accuracy to be fit for purpose. Other applications such as climate change simulations may not have the same timeliness constraints in terms of delivering a forecast, but also need to be efficient due to the long timescales for the simulations. For both of these modelling applications, the main information exchange between the surface and the atmosphere are the fluxes of heat, moisture and momentum, with the land surface acting as a bottom boundary condition. However, the increased interest in weather and climate impacts has led to additional requirements for models that represent the state of the land surface, e.g., for flooding applications, or carbon exchanges and understanding of terrestrial carbon stores in vegetation and the soil.

The complexity of the land surface schemes is also increasing as we develop our understanding of the importance of additional physical processes, such as the interaction of snow with tall vegetation (e.g., Essery et al., 2009) impacting on both the surface albedo and the timing of snowmelt, the interaction between soil moisture, vegetation and the surface fluxes of heat and moisture (e.g., Henderson-Sellers 1996, Schlosser et al., 2000), and the role of vegetation in the global carbon cycle (e.g., Cox et al., 2000). Recent insights into the interactions between the land and the atmosphere have been achieved through satellite observations (e.g., Taylor et al., 2011) and community experiments such as GLACE (The Global Land-Atmosphere Coupling Experiment, Koster et al., 2006). These studies have shown that the land/atmosphere feedbacks are stronger in certain climatic regions that in others, but that available modelling systems do not agree on the strength of coupling for these regions. Hence it is important that an urban scheme has the right level of complexity to balance required accuracy with computational cost.

1.2 Objectives of this research

The objectives of the research undertaken are:

- To identify the dominant physical processes that need to be represented within urban land surface schemes.
- To establish the complexity and parameter requirements for an urban land surface scheme to be appropriate for applications within weather and climate simulations.
- To find out if urban land surface schemes can represent the seasonal cycle of sensible and latent heat fluxes.
- To assess if the models can reproduce the observed sensitivity in the surface fluxes to

the meteorological forcing data.

- To understand the sensitivity of the urban model performance to initial conditions in order to establish the decay timescale for initial errors that influence the model results.
- To understand why initial results from the urban model comparison suggested that models that neglect the anthropogenic heat flux perform as well as models that include this term.
- To assess if the JULES model can predict the correct partitioning of energy into sensible and latent heat for urban sites with varying vegetation fractions in land cover.

Previous urban evaluation studies have been restricted to short-term observational datasets (Chapter 2), whereas in this study the performance of urban land surface models over a full seasonal cycle is investigated. This allows the impact of the level of complexity to be assessed at different times of the year when the angle of the sun, or additional warming due to human activities could be more influential.

Initial results from the comparison showed some unexpected features, such as the importance of representing vegetation and that models that neglected the anthropogenic heat flux within the energy balance had a relatively good performance. Further investigation into these aspects are undertaken to explain the results for the anthropogenic heat flux by utilising one of the urban models that contributed to the comparison. In addition, the model is also used identify if such an urban model can reproduce the influence of vegetation on the urban surface energy balance over a range or urban sites with wide variation in the vegetation fraction.

1.3 Structure of Thesis

Given the influence of the surface on the local scale climate, the impact of the urban environment and how this differs from natural surfaces is considered first (Chapter 2). This details differences in the surface energy balance, including additional terms (anthropogenic heat flux). A brief history of urban model development and observational campaigns is also provided.

The key data source for this work is the First International Urban Energy Balance Model Comparison (Grimmond et al., 2010, 2011), which included models with a wide range in complexity for various physical aspects of the urban surface energy balance. The motivation, main science goals, and design of the urban model comparison are described in Chapter 3. As the project was divided into different stages (Grimmond et al., 2011), the reasons behind this are

presented.

To address the specific objectives (section 1.2) the thesis has five focussed chapters which address these. First an analysis for the seasonal performance of the models within the comparison is presented (Chapter 4[1]). This is the first time that a larger number (21) of urban land surface models have been evaluated over the seasonal cycle within a single study. Also, the phase errors within the model simulations are considered.

Secondly, the impact of initial conditions on the performance of the models within the comparison are considered (Chapter 5[2]). Also addressed is the impact that atmospheric forcing has on model performance.

Thirdly, an overview of the results from the comparison is presented with the aim of answering the main questions posed by the comparison (Chapter 6[3]). Some differences were found compared to the original studies (Grimmond et al., 2010, 2011, Best and Grimmond, 2013, Chapter 4) with respect to the ordering of importance for the physical processes that describe the urban surface energy balance.

The remainder of the research is primarily concerned with one of the urban land surface models that took part in the comparison; the Joint UK Land Environment Simulator (JULES) model (Best et al., 2011, Appendix C[4]). This is used to investigate the impact of the anthropogenic heat flux (Chapter 7[5]), as in the comparison some models neglected the anthropogenic heat flux but were found to perform as well as those that included the term.

As representing vegetation is important for urban areas (Grimmond et al., 2010, 2011, Best and

---

[1] *This work has been published as*: Best, M.J., C.S.B. Grimmond (2013), Analysis of the seasonal cycle within the first international urban land surface model comparison, *Boundary-Layer Meteorol.*, 146, 421-446, doi: 10.1007/s10546-012-9769-7.

[2] *This work has been published as:* Best MJ, Grimmond CSB (2014) Importance of initial state and atmospheric conditions for urban land surface models performance. Urban Climate, 10, 387-406. doi:10.1016/j.uclim.2013.10.006.

[3] This work has been published as: Best MJ, Grimmond CSB (2015) Key conclusions of the first international urban land surface model comparison. Bulletin of the American Meteorol. Soc., 96, 805-819. doi:http://dx.doi.org/10.1175/BAMS-D-14-00122.1,

[4] *This work has been published as:* Best M.J., M. Pryor , D.B. Clark, G.G. Rooney, R.H.L. Essery, C.B. Ménard, J.M. Edwards, M.A. Hendry, A. Porson, N. Gedney, L.M. Mercado, S. Sitch, E. Blyth, O. Boucher, P.M. Cox, C.S.B. Grimmond, R.J. Harding (2011) The Joint UK Land Environment Simulator (JULES), Model description – Part 1: Energy and water fluxes. *Geosci Model Dev*, 4: 677-699.

[5] *This work has been published as:* Best, M.J., C.S.B. Grimmond (2015), Investigation of the impact of anthropogenic heat flux within an Urban Land Surface model and PILPS-urban, *Theoretical and Applied Climatol*. doi:10.1007/s00704-015-1554-3.

Grimmond, 2013, Chapter 4), JULES is used to investigate if it can represent the Bowen ratio for a number of urban observational datasets that vary in the fraction of vegetation cover within the footprint of the observations (Chapter 8[6]).

The final chapter (9) identifies the key contributions and conclusions of this research and recommendations for future research in terms of both observational campaigns and model developments.

In order to have an atmospheric dataset that can be used to force the urban land models, gaps in the observational time series need to be filled. The gap filling of the atmospheric forcing data used for the comparison was not included within the work undertaken for this thesis. The conducting of the comparison itself was also not undertaken for this thesis, but for chapters 4-6 all of the analysis and results form part of the contribution of this thesis, along with the authorship of the peer reviewed papers. For chapters 7 and 8, in addition to the analysis, results and authorship of the peer review papers, the contribution of this thesis also includes the appropriate developments to the JULES land surface model and all of the model simulations. Also, for the data used in chapter 8, although most of the atmospheric forcing data had already been gap filled, any additional gap filling was a further contribution to the thesis.

Three appendices provide additional work that complements the chapters of this thesis. Appendix A[7] gives initial results from Phase 1 of the comparison, whilst Appendix B[8] gives the initial results from Phase 2 of the comparison, which was for the main observational dataset. Finally Appendix C[4] gives a description of the JULES model used in Chapters 6 and 7.

---

[6] *Status of chapter*: Best, M.J., C.S.B. Grimmond, A. Christen (2015), Modelling the partitioning of turbulent fluxes at urban sites with varying vegetation cover, *submitted to J. Hydrometeorol.*

[7] *This work has been published as:* Grimmond, C.S.B., M. Blackett, M.J. Best, J. Barlow, J.-J. Baik, S.E. Belcher, S.I. Bohnenstengel, I. Calmet, F. Chen, A. Dandou, K. Fortuniak, M.L. Gouvea , R. Hamdi, M. Hendry, T. Kawai, Y. Kawamoto, H. Kondo, E.S. Krayenhoff, S.-H. Lee, T. Loridan, A. Martilli, V. Masson, S. Miao, K. Oleson, G. Pigeon, A. Porson, Y.-H. Ryu, F. Salamanca, L. Shashua-Bar, G.-J. Steeneveld, M. Trombou, J. Voogt, D. Young, N. Zhang (2010), The international urban energy balance models comparison project: first results from phase 1, *J. Appl. Meteorol. Climatol.*, 49, 1268-1292, doi: 10.1175/2010JAMC2354.1.

[8] *This work has been published as*: Grimmond, C.S.B., M. Blackett, M.J. Best, J.-J. Baik, S.E. Belcher, J. Beringer, S.I. Bohnenstengel, I. Calmet, F. Chen, A. Coutts, A. Dandou, K. Fortuniak, M.L. Gouvea , R. Hamdi, M. Hendry, M. Kanda, T. Kawai, Y. Kawamoto, H. Kondo, E.S. Krayenhoff, S.-H. Lee, T. Loridan, A. Martilli, V. Masson, S. Miao, K. Oleson, R. Ooka, G. Pigeon, A. Porson, Y.-H. Ryu, F. Salamanca, G.-J. Steeneveld, M. Trombou, J. Voogt, D. Young, N. Zhang (2011), Initial results from phase 2 of the international urban energy balance model comparison, *Int. J. Climatol.*, 30, 244-272, doi:10.1002/joc.2227

Chapter 2: The Urban Environment

2.1 Surface energy balance

The land surface has a large impact on the evolution of the near surface atmosphere from heat and moisture fluxes through turbulent exchange. The solar and longwave radiation that reaches the surface provides a source of energy for the turbulent heat and moisture fluxes (e.g., Garratt, 1992), along with a proportion of the energy that is stored in the underlying soil. The turbulent fluxes of heat and moisture are driven by gradients, between the surface and the atmosphere, of temperature and moisture respectively, whereas the surface soil flux is driven by a temperature gradient between the surface and the underlying soil. This gives equations for each of the fluxes of the form:

$$F = K\Delta S \tag{2.1}$$

where F is the flux, ΔS is the gradient in the state variable (e.g., temperature, specific humidity) and K is an effective conductivity.

For the flux into the soil, the effective conductivity is the taken to be the thermal conductivity of the soil, which will depend upon the soil properties and the amount of moisture in the soil (e.g., Best et al., 2011). For the turbulent heat flux, the effective conductivity is determined by the atmospheric turbulence and is often written in an analogous form to electrical circuits, giving for example

$$\frac{Q_H}{\rho c_p} = \frac{1}{r_A}\Delta T \tag{2.2}$$

where Q_H is the turbulent sensible heat flux, ρ is the density of air, c_p is the specific heat of air at constant pressure (giving $Q_H/[\rho c_p]$ the flux of heat), ΔT is the temperature gradient between the surface and a reference height in the atmospheric surface layer and r_A is the aerodynamic resistance.

In general, the specific humidity of the surface is not known, so typically this is represented by assuming that the surface is saturated, and then introducing an additional resistance to the aerodynamic resistance to account for a surface that is not saturated (e.g., Best et al., 2011). Again this follows the analogy of electrical resistance where the two resistances have a linear sum to give the overall resistance to moisture transport. This surface resistance is then parameterised to represent the availability of water near the surface of the soil for bare soil evaporation, or the more complex process of photosynthesis for vegetation to represent the

removal of soil water at depth through transpiration. This approach has the benefit that the saturated specific humidity depends upon the surface temperature and hence the flux equation can be written as:

$$\frac{Q_E}{L\rho} = \frac{1}{r_A + r_S}\left(q_{SAT}(T_S) - q_A\right) \tag{2.3}$$

where Q_E is the turbulent latent heat flux, L is the latent heat of condensation of water at 0 ^{o}C, q_A is the specific humidity at the atmospheric reference height, $q_{SAT}(T_S)$ is the saturated specific humidity at the surface temperature and r_s is the surface resistance.

To form a closed system, these flux equations are then combined with the surface energy balance equation (e.g., Garratt, 1992):

$$Q^* = Q_H + Q_E + Q_G \tag{2.4}$$

where Q^* is the net all wave radiation and Q_G is the flux of energy into the soil.

This equation is based upon the surface energy balance being determined over a flat plain. Whilst this is valid for surfaces such as bare soil, lakes or roads, other surfaces such as vegetation have a finite height. For these surfaces, the energy balance is actually determined over the volume that incorporates the surface type. This introduces an additional energy store for the air within this volume (ΔQ_S) which accounts for the heating and cooling of the whole volume.

For short vegetation, such as grasses, this surface volume is sufficiently small that the net storage heat flux is small compared to the other terms in the surface energy balance and hence it is often neglected . For taller vegetation, such as forests, the height of the vegetation can be tens of metres and the net storage heat flux can be a noticeable term within the surface energy balance. However, it is still sometimes neglected within surface energy balance models.

For an urban environment, not only are the buildings tall enough for the air within the surface volume to become noticeable within the surface energy balance but, unlike vegetation, the elements themselves (i.e., the buildings) have a high thermal mass which dominates the heat storage within the volume. This means that the net storage heat flux is a large term within the surface energy balance and must be represented.

Also, vegetation surfaces are usually assumed to be horizontally homogeneous with a sufficiently large fetch such that any transfer of energy due to heterogeneities, i.e., the advective heat flux (ΔQ_A), can be neglected. The heterogeneous nature of buildings within a city means

that this is not generally the case for an urban environment.

Advection in all environments is scale dependent. This term is difficult to quantify, but can be a substantial term within the energy balance equation for urban areas (e.g., Mestayer et al., 2005, Pigeon et al., 2007a). For meso-scale models the advection at the grid-scale is determined by the dynamical core of the model, but at the sub-grid-scale the advection is approximated through the choice of model parameters or neglected (e.g., Masson et al., 2002, Best et al., 2006). Urban observational experiments are often designed to minimise the advective flux (e.g., Grimmond and Oke, 1999a) so that this term can also be neglected (e.g., Coutts et al., 2007b). For example, an analysis by Steyn (1985) demonstrated that for a suburban site in Vancouver, Canada, the advection due to sea breezes was of the order of 2.5 W m^{-2} for all hours, but could be larger when individual time periods are considered (although no more than 16 W m^{-2} at its maximum).

2.2 Urban/Rural differences in the energy balance

There are a number of features for an urban surface that are substantially different from natural surfaces and lead to other adjustments in the surface flux and energy balance equations. The complex nature of the building elements within a city cause turbulent wakes that interact in various ways depending upon the spacing of the buildings (Oke, 1987). These wakes make adjustments to the atmospheric turbulence (Barlow et al., 2004, Harman et al., 2004a) that can increase the energy exchange between the surface and the atmosphere. In addition, the size and spacing of these roughness elements introduce a substantial roughness sub-layer and inertial sub-layer (e.g., Roth, 2000) which can be neglected for a rural surface.

In addition, the standard logarithmic wind profile with height does not apply within the roughness sub-layer (e.g., Macdonald, 2000). Whilst this is often neglected over rural surfaces, this region is generally of interest within a city as it includes the roughness elements, i.e., the buildings. This means that the standard surface exchange turbulence equations, that utilise an effective roughness length for the surface, can not be used in this region and so other parameterisations for the wind profile have been developed (e.g., Macdonald, 2000, Coceal and Belcher, 2004, Di Sabatino et al., 2008). However, above the roughness sub-layer the logarithmic wind profile does still hold and the standard roughness length approach is still valid, but the roughness length is altered by the turbulent wakes (Harman et al., 2004a).

To simplify the description of the built part of an urban area an urban canyon (Nunez and Oke, 1977) has often been used. The morphology of an urban street canyon is often taken to be made

up of a number of facets. These include the roofs and walls of the buildings, and the road between the buildings, although some urban models also include natural surfaces such as vegetation and bare soil. Whilst the bare soil and vegetation parts of the land cover are able to sustain a flux of moisture into the atmosphere through the water held in the soil, these urban facets are typically assumed to be unable to access moisture apart from the period following precipitation. The store of soil moisture is detached through the impervious nature of these anthropogenic surfaces, whilst runoff of excess water at the surface feeds into the drainage systems that remove this moisture from the system. This means that the turbulent moisture fluxes from such surfaces are limited to periods during which the surface can retain a small amount of water on the surface itself, and the magnitude of this flux is limited by the size of this moisture store.

Observational campaigns are rare in downtown areas of cities but more common in the more extensive residential areas (Loridan and Grimmond, 2012a). The latter can be used to investigate the impact of vegetation on the balance between turbulent heat fluxes. Grimmond and Oke (2002) showed how in general the Bowen ratio (turbulent sensible heat flux: turbulent latent heat flux) changes with increasing fraction of vegetation, based upon a number of observational campaigns. These observational studies demonstrate the importance of the moisture flux in the surface energy balance of urban areas and suggest that a representation of vegetation within urban models is important.

Another significant difference between the urban surface and those of a natural environment is the ability to store energy within the material fabric through the conductive flux from the surface. Urban building materials can have high thermal capacities and hence store large amounts of energy (e.g., Oke, 1987). In addition, the increased area from the large surface elements also increases the capability of storing energy. For natural environments, vegetation has a much lower thermal capacity and is hence not able to store such large amounts of energy. A bare soil surface can have similar thermal capacities to some building materials (e.g., Oke, 1987), but the smaller surface area and the diffusion of the energy to lower depths away from the surface change the thermal characteristics compared to the built environment. The stored energy in bare soils can be retained over longer periods of time and hence there is a lower response over the diurnal cycle.

2.3 Anthropogenic sources of heat and moisture

Human activities within cities lead to heating which needs to be taken account of within the

surface energy balance. This additional term is called the anthropogenic heat flux (Q_F). So the surface energy balance equation for an urban environment now becomes (Oke, 1988),

$$Q^* + Q_F + \Delta Q_A = Q_H + Q_E + \Delta Q_S \quad (2.5)$$

where ΔQ_S is the net storage heat flux from the surface volume, including both the urban fabric and the air within this volume.

The activities that contribute to the anthropogenic heat flux include the energy used for winter heating or summer cooling of buildings, the energy released from transportation and the heat energy generated by the metabolism of humans themselves (e.g., Sailor, 2011). Grimmond (1992) showed a diurnal cycle of these three components to the anthropogenic heat flux for a suburban area of Vancouver. In this study, the heating of the buildings was the largest term with values as large as 9 W m^{-2} and had a double peak during the morning and evening/night. The heat released through the emissions of vehicles also showed a strong diurnal cycle with values between 4 – 6 W m^{-2} during the day, but decreasing to almost zero during the night. The heat released through human metabolism showed little diurnal cycle with values during the day constant at around 1 W m^{-2}, but with almost a zero flux during the early hours of the morning. However, in the centre of some cities Q_F can be of the order of 100 W m^{-2} (e.g., Pigeon et al., 2007b, Sailor, 2011) or larger. When determined at a very fine resolution, the local anthropogenic heat flux can even exceed solar forcing at certain locations and time of day. Ichinose et al. (1999) estimated that at 25 m grid resolution centred on large high-rise office buildings in Tokyo, the hourly energy consumption could be as large as 1590 W m^{-2} in the winter and more than 400 W m^{-2} during the summer daytime hours.

This anthropogenic heat source can have both diurnal and seasonal variations (Sailor, 2011). The heating of buildings has a diurnal cycle that can be linked to the usage of the buildings. Office environments will have heating during winter days, but may not use as much energy during the evenings and night. Homes could be heated during the morning and evening, but may not be during the day if the inhabitants are located in a different place for work. Similarly, the energy released through transportation will have a diurnal cycle with peak values during the main commuter times of the working day (Sailor, 2011), whilst the location of the people in offices or at home will change the distribution of the metabolic heat source. The working behaviours of people will also determine a weekly cycle to the anthropogenic heat flux (Sailor, 2011) where there is a working week and a weekend, due to the differing transportation and building heating requirements.

The seasonal cycle in the anthropogenic heat flux is driven by the climate of the urban

environment. Cities that have cold winters are likely to have a larger energy consumption (and hence anthropogenic heat flux) during the winter months than during the warmer summers. However cities that have hot summers could have energy consumption that is dominated by summer cooling rather than the winter heating, provided that the socio-economic situation of the cities' inhabitants enables this. Changes in local climate, possibly due to climate change, are starting to impact on the seasonal cycle of energy consumption (and hence anthropogenic heat flux) with the peak energy consumption moving from winter heating to summer cooling (e.g., POST, 2008)

As well as the anthropogenic heat flux there is also an anthropogenic moisture flux. This typically takes two forms (Sailor, 2011). Air conditioning systems that remove heat from buildings can use evaporative cooling techniques, with a net impact of giving a source of moisture to the atmosphere. In addition, the combustion of hydrocarbon fuels, from either vehicles or heating systems, produce a source of moisture through chemical reactions. For instance, with vehicles, each litre of fuel burned can generate 0.9 to 1.0 kg of water vapour (Sailor, 2011).

Whilst these anthropogenic moisture fluxes give a direct source into the atmosphere, more indirect moisture sources can also be supplied to the surface. This typically takes the form of irrigation of gardens and parks in dry climates, or during dry periods (e.g., Cleugh and Oke, 1986, Grimmond et al., 1986, Grimmond, 1992), but there can also be other sources of anthropogenic moisture such as street cleaning (e.g., Oke et al., 1999). These sources of water can maintain a moisture flux that would otherwise be unsustainable, impacting on the other terms within the urban surface energy balance equation.

The surface albedo for the net shortwave radiation can be changed from that of the material elements of the building fabric due to multiple reflections caused by the morphology of the urban environment. These multiple reflections have the impact of reducing the effective albedo and increasing the net solar radiation. However, the orientation of a street canyon and the height of the buildings can lead to the shading of parts of a facet, reducing the incident solar radiation.

In addition to changes in the shortwave radiation compared to the rural surface energy balance, the tall urban structures can reduce the sky view factor (e.g., Lindberg and Grimmond, 2010). This has the impact of reducing the incoming shortwave and longwave radiation that originates from the atmospheric dome above the surface, and introduces a longwave component from the other urban facets that can be seen from the surface. This will increase the amount of incoming longwave radiation for any of the facets within the urban "surface" volume and hence is

sometimes considered as decreasing the surface emissivity (e.g., Harman and Belcher, 2006).

The increased surface area, due to the various facets of the urban morphology within the "surface" volume, also alters the nature of the area over which the surface energy balance is applied. All of these factors combine to give a complex and heterogeneous surface energy balance compared to the surrounding rural areas.

2.4 Urban observational datasets

The structure of cities is complex, even compared to surrounding rural areas. There are many physical processes that occur in urban areas, but it is not clear which are the dominant physical processes. Our basic understanding of urban environments is limited due to the many difficulties in collecting good quality observations. Compared to observational campaigns of surface fluxes in rural areas, there are a number of additional requirements for urban areas that have limited their number in comparison. For instance, finding suitable locations, power supplies and ensuring the security of the instruments themselves. The heterogeneous nature of the urban surface and the height of the roughness elements (i.e., the buildings) makes it important to ensure that the height of the observations are suitably above the roughness sub-layer. This creates an additional requirement to site instruments on tall masts and often on the top of high buildings. In addition, it is important to understand the footprint of the observations, for which the source area model of Schmid (1994) is often used.

These limitations resulted in few observational studies of surface fluxes within urban areas before the beginning of the 1990s. However during the period from 1990 to 1995 a number of campaigns were undertaken for North American cities (Grimmond and Oke, 1993, 1995, 1999a, Grimmond et al., 1996, King and Grimmond, 1997, Newton, 1999, Oke et al., 1999). These campaigns were typically of short duration, ranging from 7 to 57 days in length, and were all during the summer months with the exception of Mexico City (Oke et al., 1999) which has a sub-tropical climate. The implications are that all of these datasets were aligned to times of the year which climatologically are typically the lowest precipitation periods. With the exception of Vancouver (Grimmond and Oke, 1999a) there was just one observational site for each of the campaigns which limited the opportunity to investigate the differences between the built environment and its surroundings, or between high density down town areas and sub-urban locations.

More recent observational campaigns have tried to address the limitation of the earlier datasets.

In addition to other North American cities such as Baltimore (Crawford et al., 2011) and Oklahoma (Grimmond et al., 2004b), there have been studies in Europe for Lodz (Offerle et al., 2005a), Marseille (Grimmond et al., 2004a), Basel (Christen and Vogt, 2004), Helsinki (Vesala et al., 2008), in Africa for Ouagadougou (Offerle et al., 2005b), in Asia for Tokyo (Moriwaki and Kanda, 2004), and Australia for Preston, Melbourne (Coutts et al., 2007a, 2007b). Many of these campaigns have been for longer duration with a number of them covering a whole annual cycle (e.g., Baltimore, Helsinki, Lodz, Preston and Tokyo). In addition, some campaigns have considered a number of different sites within and surrounding the city (e.g., Basel and Oklahoma City), enabling a comparison between sites with differing urban densities.

Typically not all of the components of the energy balance have been observed. It is difficult to observe the net storage heat flux and so it is often taken to be the residual of the surface energy balance, and hence includes the observational errors from all of the other fluxes. The net advective flux is also difficult to observe and is usually neglected, although attention is often given to the location of the site to minimise the advective flux.

It is difficult to directly observe the anthropogenic heat flux, although a recent study by Kotthaus and Grimmond (2012) have attempted to calculate this term from eddy covariance measurements by detailed analysis of extreme values to identify micro-scale anthropogenic sources. Hence this term is often neglected, or derived from other information such as energy use statistics and vehicle surveys, using a method such as Sailor and Lu (2004). Similarly the anthropogenic moisture flux is not observed, with at most comments in the publications referring to irrigation (e.g., Grimmond and Oke, 1995, Grimmond et al., 1996) or street cleaning (e.g., Oke et al., 1999), or indeed irrigation bans (e.g., Grimmond and Oke, 1999a).

For some of the observational datasets, especially the earlier datasets, radiation components were not measured at the observational site. In these situations typically the shortwave radiation information comes from other nearby sites (such as airports) whilst the longwave radiation components might be determined from empirical formulae (e.g., Prata 1996).

2.5 Urban models

With the increasing requirements for the representation of urban areas within atmospheric numerical models, there are a number of urban surface schemes that have been developed in recent years (e.g., Myrup, 1969, Ackerman, 1977, Carson and Boland, 1978, Grimmond and Oke 1991, Masson, 2000, Grimmond and Oke, 2002, Martilli et al., 2002, Chen et al., 2004,

Fortuniak et al., 2004, Best, 2005, Dandou, 2005, Kondo et al., 2005, Dupont and Mestayer, 2006, Krayenhoff and Voogt, 2007, Lee and Park, 2008, Oleson et al., 2008a). These models vary in their degree of complexity by including a different number of physical processes from within the urban environment. However, increased complexity always comes at a cost. In this case that cost is in computational requirements and also in the parameter and data requirements. For operational forecasting, the weather information required by customers can not be delivered late, and so the length of the numerical model simulations is critical. Hence the computational cost of every aspect of the modelling system needs to be minimised, whilst retaining accuracy. For global applications, the requirement of parameter information for all cities that might be represented around the world can make it difficult for some complex schemes to be practical for such applications, although Masson et al. (2003) and Jackson et al. (2010) have tried to develop a database that has this capability. Hence there is a requirement to understand how all of the currently available urban schemes perform in real situations and to assess how much of the complexity of these models is required for the various applications within numerical atmospheric modelling.

The design of many urban models is based upon the street canyon as the basic modelling unit (e.g., Masson, 2000, Martilli et al., 2002, Kondo et al., 2005, Kawamoto and Ooka, 2006, Krayenhoff and Voogt, 2007, Oleson et al., 2008a, Ryu et al., 2009). As such, urban models often neglect vegetation, or indeed do not include any moisture source at all (Fortuniak, 2003, Kondo et al., 2005, Krayenhoff and Voogt, 2007). Some modelling studies have coupled such urban schemes to dry boundary layer models to study feedbacks, explicitly neglecting the influence of moisture in the feedback processes (e.g. Harman and Belcher, 2006). However, there are some urban schemes that have been coupled to a more traditional land surface model in order to include a representation of vegetation (e.g., Lemonsu et al., 2004, Chen et al., 2011), whilst others explicitly include vegetation within the urban scheme, either through the tile approach (e.g., Best, 2005, Dupont and Mestayer, 2006), or integrated into the urban energy balance (e.g., Grimmond and Oke, 1991, Grimmond and Oke, 2002, Kawamoto and Ooka, 2006, Lee and Park, 2008, Oleson et al., 2008a).

Additional complexity within the models is often introduced to include physical parts of the system which have been either neglected, or simplified to a set of parameters. However, the introduction of the new physical features will inevitably introduce a new range of parameters, resulting often in an increase in their total number. These new parameters are often linked to specific parts of the physical system, for example the properties of the building materials (e.g., Krayenhoff and Voogt, 2007). Whilst it is possible to obtain such parameters for a particular city, these parameter will vary globally (e.g., Jackson et al., 2010) which can cause difficulties

in creating the applicable parameter sets required for global modelling.

The limited availability of good quality observations within cities, the heterogeneous nature of cities from different parts of the World, or even within a city itself, and the many physical processes have made it difficult for the urban modelling community to assess which processes their models should capture. Without knowing which are the dominant processes, it is not possible to identify which processes are of second order and can be neglected for certain applications.

There have now been numerous comparisons between urban schemes and observational datasets (e.g., Grimmond and Oke, 2002, Masson et al., 2002, Best et al., 2006, Dupont and Mestayer, 2006, Hamdi and Schayes, 2007, Krayenhoff and Voogt, 2007, Kawai et al., 2009, Porson et al., 2009, Loridan et al., 2010, 2011) to assess the accuracy of the numerical urban models, with some studies using the same observational datasets (e.g., Grimmond and Oke, 2002, Masson et al., 2002, Best et al., 2006, Krayenhoff and Voogt, 2007). Typically the literature shows how the analyses have been used to improve the performance of the urban models, but do not necessarily accurately describe the initial behaviour of the model without any parameter calibration or adjustments.

However, most of the datasets are of short duration due to the complexities of observational campaigns in such environments and typically only cover periods of weeks (e.g., Grimmond et al., 1996, Oke et al., 1999, Voogt and Grimmond, 2000, Grimmond et al., 2004a, Mestayer et al., 2005, Offerle et al., 2005b, Newton et al., 2007). This means that the majority of observational campaigns have concentrated on summer periods, when the urban heat island is expected to be at its most intense. This has limited our understanding of the behaviour and the period of evaluation of urban numerical models.

2.6 Conclusions

There are many aspects of the urban surface that differ from rural areas due to human influences and hence impact on the surface energy balance. This includes impervious materials used to construct buildings and roads that alter the water retaining capabilities at the surface and the potential for stored energy through thermal mass. Whilst vegetation can often be considered to be homogeneous to a certain extent, differences in the buildings within a city mean that the urban environment is very heterogeneous. This has an additional impact when considering the energy balance at the surface through dynamical advection of the atmosphere.

Human activities also contribute to additional sources of energy through heating of buildings and vehicular transport, which should be considered in both interpreting observations and modelling. The additional sources are not limited to heat and energy, but can also include moisture sources such as irrigation or street cleaning. Availability of water has a critical impact on the surface energy balance, and hence cannot be ignored. However, these human activities are dominated by socio-economic factors as much as physical condition.

The development of urban models has been slower than that for rural surfaces because of the availability of good quality observational datasets. There are many factors that make observational campaigns more challenging in urban environments and until relatively recently these restrictions have limited observational datasets to short durations.

The implications of few observational datasets are that it has been difficult to assess the performance of urban models. Many models have been developed over the last couple of decades, with more recent models often adding new features that were not represented in previous models. This has added to the increase in complexity of these models, but the necessity of these additional features to enable a good simulation of the urban environment has not generally been demonstrated.

Chapter 3: The First International Urban Model Comparison (PILPS-Urban)

3.1 Comparison Design

In order to obtain unbiased, objective conclusions about the level of complexity required for a model to accurately represent the energy exchanges for an urban area, the first international urban model comparison was designed and implemented. This project followed the processes established with the series of successful comparison studies under the Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) (Henderson-Sellers et al., 1993, 2003). It received the support of the World Meteorological Organisation (WMO) and the Global Land Atmosphere System Study (GLASS) panel under the Global Energy and Water cycle Exchanges project (GEWEX).

The goal of the comparison was to answer the following questions:

1. What are the dominant physical processes controlling the urban surface energy balance?
2. What level of complexity is required from an urban model to be fit for purpose?
3. What are the parameter requirements for such a model and can they be determined?

To address these, the comparison was designed to investigate the simulation of various terms within the surface energy balance, in a controlled manner that enables comparable evaluation of the models. An overview of the motivation, design and goals of the comparison are given in Grimmond et al. (2009). The atmospheric forcing data for a single site was provided and for each model the user was required to return a number of essential fluxes (Table 3.1), along with the parameter information used in the simulations. Additional desirable fluxes were also requested if available.

The requirements for the evaluation data set were:

- The dataset needed to be at least twelve months in length, to ensure that the seasonal cycle of the model results could be assessed.
- The atmospheric forcing data required to run the models (Table 3.2), and the surface flux data required to evaluate the models (Table 3.1), needed to be available.
- The dataset must not have previously been used for model evaluation; so no user could have done prior “tuning” of a model for this particular site.

A wide selection of models participated in the comparison with differences in complexity and biophysical processes represented. By combining the models with similar levels of complexity for a number of physical processes, it is possible to compare the relative performance of the groups. From this, for each process, the levels of complexity required for a model to perform well in simulations can be identified. Eight categories (shown if Fig. 3.1) were identified for the comparison to represent the possible dominant physical processes, related to the way in which:

a) vegetation is represented;
b) anthropogenic heat flux is represented;
c) anthropogenic heat flux varies, either diurnally or seasonally, or not;
d) morphology of the urban area is characterised within the model;
e) facets and orientation of these facets are included;
f) number of solar reflections;
g) number of facets with a distinct albedo;
h) net storage heat flux is calculated.

Within each category there is a range of approaches from the simplest representation (e.g. neglects the process or uses a bulk approach) to the most complex representation. Each modelling group were asked to report the level of complexity in their model for each of the categories and this information is used to place the models into a number of classes from simplest to most complex (shown in Fig. 3.1). By assessing the skill of each cohort it is possible to identify what physical processes have little impact on the model results and which have the largest impact. From this analysis the dominant biophysical processes on the urban energy balance can be identified, hence answering the first of the questions.

By comparing the performance of each group of models over the various complexity classes, we can identify the level of complexity required to model the dominant physical processes. Given a particular modelling requirement, this will enable assessment of the level of complexity required for a model to be fit for this purpose. This addresses the second question.

To assess the impact of the parameter requirements, the study had multiple stages. In the first stage, no information about the site was released to the participants, beyond that it was urban. The site for these purposes was termed ‘Alpha’. The results represent the performance of the models given no specific parameter information.

At the second stage the fraction of the site surroundings that was built and vegetated was provided. Given that a large number of models use a tile scheme this enables the appropriate land cover fraction for natural surfaces to be applied. This is also information that can be

determined globally from satellites.

At the third stage, morphology data were provided (e.g., height to width ratio) which enables the various facets to be correct. This information is generally unavailable on a global scale, but could be determined using some basic assumptions and empirical relations (e.g., Bohnenstengel et al., 2011)

At the fourth stage the site location and all information about the site was provided. This included the information about building materials so participants could set up their models with the most appropriate settings. As such, we would expect the results from this stage to be the most accurate thus far.

For the final stage, the participants were given the fluxes used for the evaluation of the models. This enabled the modelling groups to optimise their models in order to minimise the errors in combinations of the fluxes. On completion of this stage, each participant was asked to return the new set of simulated fluxes along with the changes in the parameter values used for the optimisation. Analysis of the changes in these parameters enables the identification of limitations of the modelling systems through the use of parameter values that are either set outside of the observational range, or set to pre-defined limits for the parameters. This stage of the experiment took significantly more resources to complete compared to the previous stages, and as such, only a small number of participants returned results for this stage of the comparison. Due to the smaller number of models, analyses of the results have not been pursued to date.

Assessing the relative performance of the models at each stage, and taking into account the changes in information available, enables the influence of the parameters on the behaviour of the model to be determined. Hence, it is possible to identify which parameter information lead to a marked improvement in performance. This can be used to answer the third question. By studying what information is most influential, it is possible to assess if this is available either locally or globally, hence answering if the required parameter information can be determined.

If particular parameter information is identified as critical for the accurate representation of the urban energy balance, but this information is currently unavailable, it is possible to identify an essential observational requirement for the modelling community. Hence an additional goal of the model comparison experiment was to advise the observational community of the critical parameters required to represent urban environments. Although this information is not typically provided by the observational community, the results could potentially change the focus of

future observational campaigns to address the modelling community needs and hence ensure that the collected data have maximum benefit to the research community.

As with all model comparisons, it is difficult to ensure that all participants adhere to the experiment protocol (Henderson-Sellers and Pitman, *personal comm.*). To ensure maximum quality of the results in the main part of the comparison, a second site was used to obtain initial results from participants. The main focus of this phase was to ensure that the data returned met with the specifications of the comparison protocol. The site chosen was for a known location, a light industrial site in Vancouver (Grimmond and Oke, 2002), identified as VL92. This dataset provides the opportunity to assess the models at a second site, making the results from the comparison more robust.

Full information about the VL92 site was available to the participants as some modelling groups had already used these data to evaluate their models (e.g., Masson et al., 2002, Best et al., 2006, Oleson et al., 2008a). Hence all modelling groups were able to develop/evaluate their models with these data. Thus, the VL92 results can not be used to assess the impact of parameter information, but can be used to identify the dominant physical processes and level of complexity required for the models to be fit for purpose. As this dataset covered only a short period it was not possible to assess the seasonal performance.

3.2 The VL92 dataset

The Phase 1 site was a light industrial area in Vancouver, Canada. The site consisted of one- and two-storey warehouse structures with virtually no vegetation (Grimmond and Oke, 1999a), and can be characterised by the Urban Terrain Zones (UTZ) of Ellefsen (1985) as Do4. The mean building height was 5.8 m ± 0.1 m, the mean height of the trees 8.6 m ± 0.2 m, which gave a mean height for the roughness elements of 6.0 m ± 0.2 m averaged on plan area and 6.7 m ± 0.1 m averaged on frontal area (Grimmond and Oke, 1999b). The observation period was from the 10th August 1992 to the 25th August 1992 (day of year: 223 – 238).

Geographic Information Systems (GIS) which contain the surface descriptors were developed from areal photographs and field surveys at the site, following the procedures defined in Grimmond and Souch (1994) and Grimmond (1996). The source areas for the hourly measured convective fluxes were determined using the Flux Source Area Model (FSAM) of Schmid (1994). The source areas were calculated assuming a non-dimensional height $z_{ref}/z_0 = 33$ (where z_{ref} is the reference height of the sensors and z_0 is the roughness length for momentum), near-

neutral stability (z_{ref}/L = -0.04, where L is the Obukhov length) and a lateral turbulence parameter σ_v/u_* = 1.9 (where σ_v is the horizontal crosswind standard deviation of the wind fluctuations and u_* is the friction velocity). The source areas contributing to 90% of the concentration of the flux were used for the calculations. This analysis determined area surface cover fractions of 4% for vegetation, 32% for impervious ground, 37% roofs and 27% walls.

At the observational site, the instruments were mounted on tall towers to ensure that the measurements were within the constant flux layer of the urban boundary layer. This means that the measurements are representative of the local-scale with horizontal length scales of $10^2 – 10^4$ m (Grimmond and Oke, 1999a). The net all-wave radiation was measured using either a Swissteco or REBS net pyrradiometer. Vertical wind velocity fluctuations were measured using a Campbell Scientific Inc (CSI) one-dimensional sonic anemometer and used, along with a fine wire thermocouple system for temperature fluctuations and a CSI krypton hygrometer for absolute humidity fluctuations, to determine turbulent sensible and latent heat fluxes using the eddy correlation method. These fast response instruments were mounted less than 0.15 m apart and the fluctuations were sampled at 5 or 10 Hz and the covariances were averaged over 15 minute periods. Flux corrections were made for oxygen absorption and air density, but no corrections were applied for frequency response or spatial separation of the sensors.

The storage heat flux was determined from the residual of the basic surface energy balance equation:

$$\Delta Q_S = Q^* - \left(Q_H + Q_E\right) \tag{3.1}$$

This means that all observational errors in the other fluxes are accumulated into the net storage heat flux. For all of the fluxes, these errors can include those due to the spatial inconsistency of the energy balances, which result from variations in the source area due to fluctuations in the wind direction, atmospheric stability and surface roughness. For the radiant fluxes, these surface characteristics remain fixed in time except for shadows (Soux et al., 2004)

For this observational study, the anthropogenic heat flux and the horizontal advection terms within the urban energy balance (eqn. 2.5) were not measured. The anthropogenic heat flux enters the climate system in a number of ways. Warmer surfaces due to this flux will act as an additional source of radiation and additional convected sensible heat (Grimmond and Oke, 1999a). Additional sensible and latent heat flux will also be convected directly from the exhaust pipes of vehicles and chimneys or leakage from buildings. However, some sensible heat will be conducted into net storage heat flux. The radiative and convective components from the anthropogenic heat flux will be sensed by the instruments, so only the storage component in the

fabric of the buildings will be in the net storage heat flux (Grimmond and Oke, 1999a). This highlights that the basic surface energy balance equation used to determine the net storage heat flux as the residual is a measured energy balance. A numerical model of this environment needs to explicitly include the anthropogenic heat flux term in its energy balance equation.

It is difficult to assess the impact of the advection flux on the errors in the net storage heat flux. There will have been some mesoscale circulations that could contribute to an advective flux, however the site was selected to minimise the impact. This means that the advective flux was probably small and was hence neglected.

Although Vancouver usually experiences a mild summer drought, typically a surface moisture source is maintained from considerable sprinkler irrigation. This is typical of other cities, as discussed in Chapter 8. However, the summer of 1992 was drier than usual which led to a ban on external irrigation (Grimmond and Oke, 2002). This means that there was little surface moisture available for evaporation during the observational period and hence only a small latent heat flux. The implication of this is that the impact of neglecting moisture or vegetation within urban numerical models should be small.

3.3 The "Alpha" dataset

The "Alpha" dataset was collected in Preston, a suburb in north Melbourne, Australia (Coutts et al., 2007a, 2007b). The area consists of moderately developed low-density housing and is defined as being within the Urban Climate Zone (UCZ) 5 (Oke, 2004). The mean height to width ratio of the site is 0.42 and the mean wall-to-plan ratio is 0.40 (Coutts et al., 2007b). Aerial photographs for 2002 used for surface cover estimates, give the fraction of pervious surface area as 38%, of which 23% was defined to be dense vegetation (Coutts et al., 2007a), although the form and type of vegetation was not provided. These fractions consist of impervious as: buildings (44.5%), concrete (4.5%), and roads (13%); and pervious: vegetation excluding grass (22.5%), grass (15%) and other (bare or pools, 0.5%).

Radiative fluxes were measured at the site using Kipp and Zonen CM7B and CG4 radiometers. The data were sampled at 1 Hz and averaged to 30 minutes. Wind velocities were measured using a CSI CSAT 3D sonic anemometer whilst humidity was measured using a CSI KH20 krypton hygrometer during the period from August 2003 to February 2004, and a LICOR LI7500 open path infrared gas analyser for the remaining period. These were sampled at 10 Hz and used to calculate the turbulent heat and moisture fluxes at 30 minute intervals using eddy

covariance techniques.

The anthropogenic heat flux was estimated for this site following the method of Sailor and Lu (2004), utilising locally available data. However, only a simplified version of this method was adopted because detailed databases such as those used by Sailor and Lu (2004) were not available for Melbourne. Components of this heat source were determined for three main sources: vehicles, building sector and human metabolism. The heat flux from vehicles was determined with data from a vehicle survey during November 2002 – October 2003. For the heat released from buildings, data on the half-hourly demand for electricity and the daily consumption for natural gas was utilised. The human metabolic rate was assumed to take different values for night-time hours, daytime hours and transitional periods with values of 75 W, 175 W and 125 W respectively, with the anthropogenic heat flux for human metabolism then determined from the population density. The total anthropogenic heat flux was taken as the sum these three components (see Coutts et al., 2007b, for more detail).

The urban net storage heat flux was not measured directly, but was determined as the residual of the urban energy balance (eqn 2.5), assuming that the advective flux is negligible and can hence be neglected. The disadvantage of this approach is that the net storage heat flux accumulates the observational errors from the other terms within the energy balance equation. Coutts et al. (2007b) suggest that the sensible and latent heat fluxes could be under-estimated by as much as 20%, so the residual value of the net storage heat flux should be considered as an upper limit of the energy stored. The net advection flux is assumed to be negligible in the observations based on appropriate siting of the instrumentation to have extensive fetch that is regarded as being homogeneous.

The data were collected over a 474 day period, although not all of the data were available at all times. For the forcing data for the urban surface models, the missing data were gap filled in order to obtain a continuous forcing dataset with a half hour time-step. For the surface fluxes, which are used to evaluate the urban models, the missing data have not been gap filled, but the analysis is only undertaken for time periods where all of the fluxes are available. This limits the analysis to 8865 intervals (38.9% of the total time period), but allows for a consistent comparison between the fluxes in the surface energy balance.

3.4 Initial analyses of results

Results from 33 models were returned for the comparison, although one model did not complete the runs for stage 4 and hence was included in the analysis of VL92 dataset (Grimmond et al., 2010, Appendix A), but was excluded from the subsequent analyses on the "Alpha" dataset (Grimmond et al., 2011, Appendix B and Chapters 4, 5 and 6). Of the 32 remaining sets of model simulations, these were performed by 24 different users with 21 independent models. A list of the models is given in Grimmond et al. (2010, Appendix A, 2011, Appendix B) and Chapters 4, 5 and 6.

A number of statistical metrics were used to assess the performance of the models for each flux compared to the observations. These were: mean ($\bar{x}$), standard deviation (σ), Pearson's correlation coefficient (R) and coefficient of determination (R^2), root mean square error (*RMSE*), systematic *RMSE* (*RMSEs*), unsystematic *RMSE* (*RMSEu*), mean absolute error (*MAE*), mean bias error (*MBE*) and index of agreement (d). The definition for these statistical metrics is given in Grimmond et al. (2010, Appendix A). However, the analyses presented in Chapters 4, 5 and 6 consider only the metrics for mean, MBE and RMSE.

Initial results from stage 1 of the comparison experiment using the VL92 dataset are given in Grimmond et al. (2010, Appendix A). This study concluded that the models performed best at predicting the net radiation flux and were worst with the latent heat flux. No model was best at predicting all of the fluxes, and it is difficult for the models to minimise errors in both the net radiation flux and the sensible heat flux. Generally the unsystematic errors were smaller than the systematic errors, especially for the poorest performing models.

Models that do not represent vegetation gave the poorest performance for all fluxes over both the day and night, even at this site which was characterised with limited vegetation fraction. During the day, models that do not include the anthropogenic heat flux do at least as well as the models that do include this term in the surface energy balance equation. However, during the night, the models that include the anthropogenic heat flux have better performance for the net radiation and the net storage heat flux.

Initial results from the comparison experiment using the "Alpha" dataset are presented in Grimmond et al. (2011, Appendix B). Again, it was concluded from this study that no individual model performed best over all of the evaluation fluxes. Implications of this could be that some models have a good representation of one flux, but for the wrong physical reasons. Representing vegetation significantly impacted on the performance of the models, which agreed with the

results using the VL92 dataset even though the later had a much lower plan area fraction of vegetation. Indeed, specifying the fraction of surface cover had the largest impact on performance for most of the models. In addition, the *RMSE* in the latent heat flux was of the same order of magnitude as the observed flux, showing that there needs to be a focus on the way in which moisture exchange is represented.

The study also concluded that closure of the radiation balance is not a good indicator of the ability of a model to represent a particular radiative flux. Whilst the upward component of the shortwave radiation was generally modelled well, the upward longwave component was not as good, although the net radiation is modelled better than either of the two radiative components. Models with the simplest characteristics tended to out-perform the more complex models as they could assign a single parameter value that matched the observations. These simple models showed a net improvement with the additional site information given at each stage, whereas the more complex models did not. However, using an ensemble of models generally out performed a single model.

3.5 Conclusions

The urban model comparison was the first international comparison experiment undertaken by the urban community. The aims of the experiment were to compare participating models within a consistent framework that enables an objective assessment. As such, the experiment was carefully designed in stages to examine various aspects of the models.

The applications for urban models are varied, so the comparison concentrated on a single application, that of weather forecasting and climate prediction. The requirements of an urban model for this application are to have accurate surface fluxes of net radiation and sensible and latent heat, which form boundary conditions to the atmospheric model.

There was good participation within the comparison from a number of models and groups. This not only enabled comparison between the models, but also a comparison of the same model run by different groups, so that model assumptions could be studied.

Initial analyses of the results from the comparison showed that no model performs best, which could suggest that none of the models have an optimal configuration. In addition, representing vegetation is critical to get the correct balance between heating and moistening of the atmospheric boundary layer over an urban environment.

Table 3.1: Fluxes used for the evaluation of the models in the comparison

Flux (W m^{-2})	Essential/Desirable
Net all wave radiation flux	Essential
Sensible heat flux	Essential
Latent heat flux	Essential
Net storage heat flux	Essential
Outgoing shortwave radiation flux	Desirable
Outgoing longwave radiation flux	Desirable

Table 3.2: Atmospheric forcing data required for the model comparison

Atmospheric forcing	Units
Incoming shortwave radiation	W m^{-2}
Incoming longwave radiation	W m^{-2}
Rainfall	kg m^{-2} s^{-1}
Air temperature	K
Relative humidity	%
Wind speed	m s^{-1}
Wind direction	^{o}
Atmospheric pressure	Pa

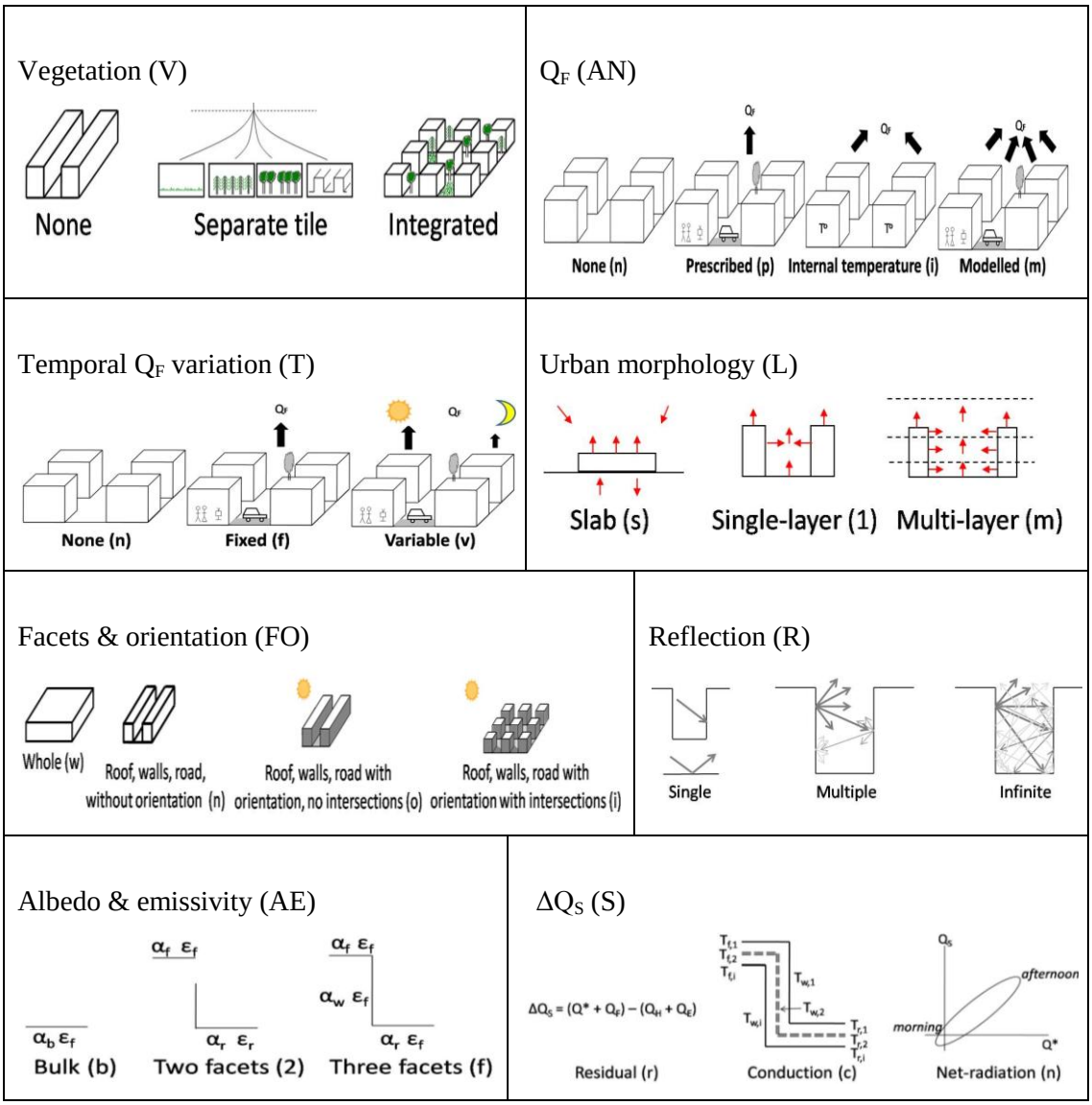

Figure 3.1: Categories based upon possible dominant physical processes for an urban area, that are used to group models in the urban model comparison. The classes within each physical category are based upon the complexity of the processes represented, ranging from the simplest to the most complex. (Adapted from Grimmond et al., 2010).

Chapter 4: Analysis of the Seasonal Cycle within the First International Urban Land Surface Model Comparison [9]

Abstract

A number of urban land-surface models have been developed in recent years to satisfy the growing requirements for urban weather and climate interactions and prediction. These models vary considerably in their complexity and the processes that they represent. Although the models have been evaluated, the observational datasets have typically been of short duration and so not suitable to assess performance over the seasonal cycle. The first international urban-land surface model comparison used an observational dataset that spanned a period greater than a year, which enables an analysis over the seasonal cycle, whilst the variety of models that took part in the comparison allows the analysis to include a full range of model complexity. The results show that, in general, urban models do capture the seasonal cycle for each of the surface fluxes, but have larger errors in the summer months than in the winter. The net all-wave radiation has the smallest errors at all times of the year but with a negative bias. The latent heat flux and the net storage heat flux also are underestimated, whereas the sensible heat flux generally has a positive bias throughout the seasonal cycle. A representation of vegetation is a necessary, but not sufficient, condition for modelling the latent heat flux and associated sensible heat flux at all times of the year. Models that include a temporal variation in anthropogenic heat flux show some increased skill in the sensible heat flux at night during the winter, although their daytime values are consistently overestimated at all times of the year. Models that use the net all-wave radiation to determine the net storage heat flux have the best agreement with observed values of this flux during the daytime in summer, but perform worse during the winter months. The latter could result from a bias of summer periods in the observational datasets used to derive the relations between net all-wave radiation and the net storage heat flux. Apart from these models, all of the other model categories considered in the analysis result in a mean net storage heat flux that is close to zero throughout the seasonal cycle, which is not seen in the observations and could imply that forcing energy balance closure through the net storage heat flux at short timescales is not appropriate. Models with a simple treatment of the physical processes generally perform at least as well as models with greater complexity.

---

[9] *This has been published as:* Best, M.J., C.S.B. Grimmond (2013), Analysis of the seasonal cycle within the first international urban land surface model comparison, *Boundary-Layer Meteorol.,* 146, 421-446, doi: 10.1007/s10546-012-9769-7.

4.1 Introduction

With the majority of the world's population now living in urban areas, there is an increasing requirement for accurate forecasting of weather conditions within the urban environment. As the built environment is only a small percentage of the land surface, from a global perspective, numerical weather forecasting and climate prediction models traditionally have not resolved these areas and hence neglected the impact of both the urban areas on the evolution of the atmosphere and the change in meteorological conditions within the city compared to surrounding rural areas (Best 2006). However, advances in computing power have enabled the resolution of such models to increase to the extent that cities can now be resolved in the high resolution limited area models, and even within the global models for some of the larger cities in the World.

The latest generation of land-surface models have adopted the tile, or mosaic, approach to represent the heterogeneity of the land surface within a numerical grid box (e.g., Avissar and Pielke 1989; Koster and Suarez 1992; Essery et al. 2003). This means that these schemes calculate the surface energy balance for a number of different surface types and blend the derived surface fluxes at an appropriate height. As such, numerical models are now capable of representing some of the impacts of the urban areas, even if the cities are not explicitly resolved.

With these increasing requirements for the representation of urban areas within atmospheric numerical models, there are a number of urban surface schemes that have been developed in recent years (e.g., Table 4.1). The models vary in their degree of complexity by including a different number of physical processes from within the urban environment. However, increased complexity always comes at a cost, mainly in terms of computational needs and also in the parameter and data requirements. For operational meteorology, the timeliness of the forecast information is imperative, and so the computational cost of any scheme becomes important and there is a requirement to minimize this cost whilst maintaining accuracy. For global applications, the parameter information needed to represent all cities make it difficult for some complex schemes to have all the complete parameter information. However, Jackson et al. (2010) recently developed a database to address some aspects of this by region and urban density, and Loridan and Grimmond (2012a) took a simplifying approach to determine how many urban zones are needed for energy partitioning (UZE) and to assign parameter values (Loridan and Grimmond 2012b). There is a need to understand how the currently available urban land-surface models perform in real situations and to assess how much of the complexity of these models is required for the various applications within numerical atmospheric modelling.

Despite numerous comparisons between urban land-surface models and observational datasets (e.g., Grimmond and Oke 2002; Masson et al. 2002; Dupont and Mestayer 2006; Hamdi and Schayes 2007; Krayenhoff and Voogt 2007; Kawai et al. 2009; Porson et al. 2009; Loridan et al. 2010, 2011) to assess their performance, these have typically been for short duration (few weeks) due to the complexities of observational campaigns in cities. Compared to rural campaigns, the additional issues of instrument security, permissions and access to suitable sites, and access to power supplies, have meant the majority of experimental campaigns have concentrated on summer periods (e.g., Grimmond et al. 1996; Oke et al. 1999; Voogt and Grimmond 2000; Grimmond et al. 2004a; Mestayer et al. 2005; Offerle et al. 2005b; Newton et al. 2007), when the surface fluxes are expected to be largest. This has limited our understanding of the behaviour and the period of evaluation of urban numerical models.

Until the international Project for Intercomparison of Land surface Parameterization Schemes (PILPS, Henderson-Sellers et al. 1993, 2003, Irranejad et al. 2003) with urban energy balance models (PILPS-Urban, Grimmond et al. 2010, 2011), schemes had been evaluated using the same datasets (e.g., Grimmond and Oke 2002; Masson et al. 2002; Best et al. 2006; Krayenhoff and Voogt 2007; Oleson et al. 2008a), but lacked a consistent methodology. For PILPS-Urban, an observational dataset was selected that spanned more than one year, enabling the analysis of schemes over the seasonal cycle and the relative merits of differing complexity within these models during different conditions over the year; for example, the impact of vegetation within urban areas over differing seasons.

The aim of the comparison was to assess the complexity required to represent urban areas within a numerical atmospheric model. As such, no attempt was made during the collection of the comparison data, or with subsequent analyses, to rank the models in an identifiable way or to name the models that performed best or worst. The analysis has simply been applied to identify common physical characteristics that give relatively ‘good’ or ‘poor’ performance. In this paper the same ethos of anonymity for the modelling groups is adopted.

The factors that influence the seasonal cycle of the surface fluxes include the interaction of the solar radiation with the urban surface, both in terms of the magnitude of the solar flux and also the zenith angle, leading to a different fraction of shading on various surfaces. Anthropogenic activities can have a substantial seasonal cycle. In many climates, winter heating dominates and leads to a larger anthropogenic heat flux in winter than in summer, although this is not solely determined by climate but also depends upon socio-economic conditions. However, where air conditioning is common there are now areas where anthropogenic heat flux from summer

cooling exceeds that from winter heating, especially where there is a mild winter. The diurnal and seasonal pattern in the movement of humans around the urban environment (e.g., changing population densities during working hours) can make a contribution to variations in the anthropogenic heat source through metabolic heat generation and transportation sources.

It is well known that the seasonal cycle of soil moisture can affect the water availability for evapotranspiration by vegetation (e.g., Dirmeyer et al. 2006; Notaro et al. 2010). During summer the available energy from the radiative fluxes is relatively larger than in winter, but dry periods limit evaporation compared to periods of plentiful water supply. On the other hand, winter periods that are unrestricted by soil moisture have evaporation limited by the available energy. In addition, anthropogenic moisture sources can enhance soil moisture in dry periods through irrigation leading to increased evapotranspiration, or direct evaporation can occur from water used for street cleaning, etc. So to understand the full impact of vegetation within an urban area, the seasonal cycle needs to be studied.

The objective of our study reported herein is to assess the relative merits of the varying complexities of urban land surface models and to identify the dominant physical processes required to represent the annual variations in the surface fluxes. This is done through an analysis of the seasonal behaviour from the first international urban land surface model comparison.

4.2. Methods

4.2.1 Model Characteristics

To evaluate seasonal performance, 32 models that participated in the comparison study (Table 4.1) were categorized based on the physics used, following Grimmond et al. (2010, 2011) (Table 4.2): (i) include a representation of vegetation, (ii) temporal variation in the anthropogenic heat source, (iii) facets and orientation of the urban structure, (iv) number of solar reflections within a street canyon, and (v) number of facets with distinct albedos.

The three approaches to vegetation are termed (in order of increasing complexity, Table 4.2): none, tile, and integrated. For the tile approach, the surface energy balance is calculated separately for the vegetation and urban fractions, and the surfaces only interact through the evolution of the atmospheric variables (e.g., temperature) so there is no interaction during a timestep. In contrast in the integrated approach, the vegetation surface fluxes are influenced by the other surfaces (and vice versa) through updating the urban canopy layer temperature and

humidity through which each of the surface fluxes depends, leading to an interaction within the timestep.

The additional energy supplied by anthropogenic heat has distinct diurnal and seasonal cycles (Allen et al. 2011; Sailor 2011) because of human activities (e.g., vehicle use and population density changes) and regional climate and socio-economic factors (e.g., heating or cooling of buildings). These temporal changes in the anthropogenic heat flux may be expected to affect the surface energy exchanges. So the treatment of the anthropogenic heat flux within a model could have an impact on the seasonal behaviour of the calculated surface fluxes.

The zenith angle and related magnitude of the incoming solar radiation varies through the year. These two factors can have complex interactions within an urban environment. Low zenith angles are associated with weaker incoming radiation and can lead to shading on some of the facets of the urban environment. However, due to the low solar angle, there may be many reflections from the facets that are sunlit within the complex urban structures, increasing the fraction of energy that is absorbed. In contrast, high zenith angles with associated greater incoming solar radiation, give rise to a small fraction of the facets being shaded, but may have fewer reflections and hence absorb a smaller percentage of the solar energy. The net result of these interactions can depend upon the orientation of a street canyon, with the potential for more shading of incident solar radiation that is normal to the street canyon compared to a canyon that is aligned to the sun's rays. Also, the maximum number of reflections considered by the urban models will have an impact, as will the number of distinct facets that are represented with a unique albedo. Hence the approach for representing the orientation of a street canyon, the number of reflections and the number of facets with a distinct albedo are all likely to have an impact on the seasonal behaviour of the models.

By combining each of the categories describing the physics of the models, we can define a category for the overall complexity of the models. Following Grimmond et al. (2011), a bimodal classification (simple or complex) for each category can be identified and the overall complexity can then be determined by the total number of simple classifications: complex models have no simple classifications; medium complexity models have one or two simple classifications; and simple models have more than two simple classifications. Grimmond et al. (2011) showed that a representation of vegetation had a large impact on the resultant negative behaviour of the models, and hence the vegetation category is omitted when defining the overall complexity of the models, so as not to unduly influence the results. This overall complexity classification helps to identify the level of complexity required to resolve the observed seasonal cycles in the fluxes.

4.2.2 Model runs and analysis methods

The analysis is performed using the model runs from Phase 2/Stage 4, the main phase of the comparison, and the stage when all information about the site characteristics (Sect. 4.2.3) was released to the participants. The impact of parameter selection and requirements on the performance is not studied as this will be the focus of a future paper. However, it should be noted that Grimmond et al. (2011) found that not all models improved over each of the comparison stages, implying that some models (or modellers) are not able to correctly use the local information provided.

For Stage 4 there are 32 model runs that can be analysed, and consists of 21 independent models run by 24 different users (Table 4.1). Some groups submitted more than one set of results for an individual model (e.g., using different versions of their model) and some models were run by more than one group.

To assess the seasonal behaviour of the models within each of the categories, the modelled fluxes are compared to observations for each model over 60-day periods (Sect. 2.3). The mean bias error (MBE) and root mean square error (RMSE) statistics are generated independently for each model and for each 60-day period to generate seasonal values. For each class category the median is calculated to ensure that outlying model results do not bias the category results. This means that in the seasonal time series there can be a switch between models through the year.

4.2.3 Observational data and site

The forcing and evaluation data were collected in Preston, a northern suburb of Melbourne, Australia (Coutts et al., 2007a, 2007b) for the period August 2003 to November 2004. The moderately developed low-density housing is classified as an Oke (2004) urban climate zone (UCZ) 5 (Coutts et al. 2007b) or a Loridan and Grimmond (2012b) urban zones to characterise energy partitioning (UZE) medium density (MD). The mean height-to-width ratio of the site is 0.42 and the mean wall-to-plan ratio is 0.40 (Coutts et al. 2007b). Analysis of 2002 aerial photographs characterize the surface cover as 38% pervious, with 23% dense vegetation (Coutts et al. 2007a), but the form and type of vegetation are not provided. The impervious area consists of buildings (44.5%), concrete (4.5%), and roads (13%); and pervious: vegetation excluding grass (22.5%), grass (15%) and other (bare or pools, 0.5%).

Radiative fluxes were measured on the site at a height of 40 m using Kipp and Zonen CM7B and CG4 radiometers, with the data sampled at 1 Hz and averaged to 30-min. Wind speeds were measured using a CSI CSAT 3D sonic anemometer whilst humidity was measured using a CSI KH20 krypton hygrometer during the period from August 2003 to February 2004, and a LICOR LI7500 open-path infrared gas analyser for the remaining period. The 10 Hz samples were used to calculate the turbulent heat and moisture fluxes at 30-min intervals using eddy-covariance techniques.

The anthropogenic heat flux was estimated by Coutts et al. (2007b) following the method of Sailor and Lu (2004) for vehicles, building sector and human metabolism. This includes using: a vehicle survey during November 2002 – October 2003; 30-min electricity and daily natural gas data; and nighttime, daytime and transitional period metabolic rates (75 W, 175 W and 125 W respectively), with population density to determine the total anthropogenic heat flux.

The net storage heat flux was determined as the residual of the urban energy balance. This assumes that the advective flux is negligible, based on appropriate siting of the instrumentation with extensive fetch that is regarded as homogeneous. The disadvantage of this approach is that this flux accumulates the observational errors from the other terms within the energy balance equation. Coutts et al. (2007b) suggest that the sensible and latent heat fluxes could be under estimated by as much as 20%, so the residual value of the net storage heat flux should be considered as an upper limit of the energy stored/released.

The data were collected over a 474 day period (13 August 2003 to 28 November 2004). For forcing requirements, missing data were gap filled in order to obtain a continuous dataset with a 30-min timestep (Grimmond et al. 2011). To evaluate the models, analysis is only undertaken for time periods when all the fluxes are available. This limits the analysis to 8865 intervals (38.9% of the total time period), but allows for a consistent comparison between the fluxes in the surface energy balance. The analysis is conducted for seven 60-day periods starting after the first 54 days, which are treated as a spin-up period for the models.

4.3. Results

The site is characterized with observed daytime turbulent sensible heat (Q_H) and net storage heat (ΔQ_S) fluxes that are of similar magnitude throughout the year compared to the smaller latent heat flux (Q_E) (Fig. 4.1). Unlike the daytime rural soil heat flux, which is typically a small proportion of the net all-wave radiation (Q^*), ΔQ_S plays a more dominant role in urban areas.

During the summer months, Q_H generally has larger values than ΔQ_S, but this is reversed in winter months (Fig. 4.1). Coutts et al. (2007a, 2007b) provide more detailed discussion of their observed fluxes.

Generally, the diurnal cycles are captured by the models throughout the year (e.g., Fig. 4.1) with the biases in the mean flux dominated by the daytime bias in each of the fluxes. This is not surprising as this is the time of day with the largest fluxes. Another general feature of the models' performance is a phase shift in the diurnal cycle of ΔQ_S. In the morning hours the observed and modelled fluxes increase concurrently but the peak and subsequent decrease occur too in early in the models, which leads to larger negative biases in the early afternoon period (e.g., Fig. 4.1).

In general, considering the results across all categories (Table 4.2), the models are able to capture the seasonal variations in each of the mean fluxes (e.g., Fig. 4.2), although there is a seasonal cycle to the *RMSE* that follows the absolute values of the mean fluxes (Fig. 4.2). However, if the *RMSE* is normalized with the absolute value of the mean observed flux for each 60-day period, then we find that the normalized errors are quite consistent throughout the year (e.g., Fig. 4.3). The peak in the winter June/July period are a result of the small observed absolute mean values for each of the fluxes, with the exception of the latent heat flux, with magnitudes less than 10 W m^{-2} (Fig. 4.2). The other larger values for the normalized *RMSE* with the sensible heat flux are also due to small absolute values for the mean observed flux. Hence only the net storage heat flux shows a real change in the normalized *RMSE* throughout the seasonal cycle, with an increase occurring from the summer through to the winter period.

Of the four fluxes considered, Q^* has the smallest errors at all times of the year. In general there is a negative bias through the seasonal cycle for Q^*, as there is for Q_E and ΔQ_S, whereas Q_H generally has a positive bias throughout the year (Fig. 4.2).

4.3.1 Representation Of Vegetation

In general models that do not explicitly include vegetation are not able to reproduce Q_E, leading, on average, to no moisture flux from the urban area at all (Fig. 4.1 and 4.2). For Q^* and ΔQ_S there is little difference between the models that do and do not include vegetation, with all three categories showing a similar seasonal pattern in these fluxes (Fig. 4.2). The models that neglect vegetation compensate for the lack of Q_E by having Q_H that is too large during all times of the year. The mean bias errors for Q_H and Q_E are close to mirror images for the models that do not

include vegetation (Fig. 4.2). For both fluxes there are larger biases in summer than winter, as would expected given the larger summertime flux values. This is confirmed by normalizing the biases by the absolute value of the mean observed flux (Fig. 4.3), which gives a consistent signal throughout the seasonal cycle for all fluxes, apart from the winter (June/July) when the mean observed fluxes are small. However, Q_H has a larger *MBE* throughout the year than Q_E (Fig. 4.2). This additional offset in the mean fluxes is largely compensated by a bias in ΔQ_S. Unlike Q_H and Q_E, the *MBE* for Q^* and ΔQ_S does not show a large seasonal cycle (Fig. 4.2).

The models that do not include vegetation clearly have the largest *RMSE* for Q_H and Q_E compared to the other model classes (Fig. 4.2). For summer, the *RMSE* for both Q_H and Q_E is larger than the observed mean fluxes, demonstrating that neglecting vegetation has a detrimental impact on modelling these fluxes. Despite this being the time of year when water at the surface may be limited, the maximum solar forcing ensures there is no limitation in available energy, so vegetation continues to access soil water that maintains evaporation, although it is smaller than typical rural energy balance fluxes. However, in the Melbourne context because of drought conditions, the surrounding rural vegetation may have had even smaller fluxes. There are periods when irrigation does occur in the suburban areas but this is also subject to restrictions and bans.

In summer, models that neglect vegetation have a larger *RMSE* (unlike the *MBE*) for Q^* than the other classes (Fig. 4.2). In winter the class of models that neglect vegetation only has a low *RMSE* for ΔQ_S. The two classes that include vegetation generally have little difference in their performance (Fig. 4.2). The model cohort that uses the tile approach to vegetation has the smallest *MBE* in Q_H at most times of the year but larger *RMSE* than the integrated vegetation cohort. The tile schemes have the smaller *RMSE* for ΔQ_S whereas integrated models have the smaller for Q^* (Fig. 4.2).

It is clear from the diurnal cycle of Q_E (Fig. 4.1) that there is no evaporation from most of the models that do not include vegetation. The impact on Q_H is a large overestimation of the daytime peak value during the summer and a smaller, but still substantial, overestimation of the peak values in the winter. The nocturnal Q_H is also larger than for the other two categories; in summer this is in better agreement with the observations, but is an overestimate in winter (Fig. 4.1). The models that lack vegetation, not unexpectedly, have larger daytime ΔQ_S, which actually results in better agreement with the observations in general than is the case with other classes (Fig. 4.1).

The models without vegetation have larger daytime peak Q^* values than the other two classes

and the observations (Fig. 4.1). This is unexpected, since without an evaporative flux, the surface temperature is likely to be larger (consistent with larger Q_H and ΔQ_S), resulting in increased outgoing longwave radiation. This would act to reduce Q^* and so the resultant increase must be larger than the changes to the outgoing longwave component. As the incoming radiation is a specified forcing in the comparison, this implies that, on average, a lower effective albedo is being used. As the effective albedo for urban surfaces is typically lower than for natural surfaces, it is possible that the increased urban land-cover fraction (due to the neglect of vegetation) has the impact of reducing the effective albedo compared to the other classes.

It is not clear whether urban vegetation should have the same effective albedo as similar vegetation in a rural environment. Multiple reflections of radiation due to the morphology of the urban area could lead to an increased absorption by the vegetation, resulting in a lower effective albedo. However, albedo values that are the same as rural vegetation is typically the assumption made for models with a tile approach. The average diurnal cycle for Q^* (Fig. 4.1) from tile models has smaller peak values during the day than observed throughout the year, especially in the summer. The integrated vegetation models, despite having larger peak values than the tile approach, still underestimate the observed peak (Fig. 4.1). This may suggest that the urban vegetation albedo should be a function of the surrounding urban morphology and not just assigned a rural value.

As indicated, the differences between the two vegetated classes are much smaller than the differences between including and not including vegetation. Over most of the year the integrated vegetation models tend to have a larger peak Q_E than occurs in the tile approach, except December/January, during which period there is an increased ΔQ_S that is similar in magnitude to the models that neglect vegetation (Fig. 4.1). It is unclear whether a reduced Q_E increases ΔQ_S, or an increased ΔQ_S leads to a reduced Q_E. However, one possibility is that the additional evaporation over the October/November period reduces the soil moisture to a point where the evaporation from the vegetation starts to become moisture limited. Such a situation would increase the sensible heat flux from the vegetation leading to higher canopy-space air temperatures. This would reduce the temperature gradient between the urban surface and the canopy space suppressing Q_H from this surface, which would result in an increased ΔQ_S, since there is no evaporation directly from the urban surface itself, except after periods of precipitation. This is not possible in the tile approach because this interaction would have to occur indirectly through the atmospheric temperature, but these are specified in the offline comparison and hence this feedback is removed.

4.3.2 Temporal Variation in Anthropogenic Heat Flux

The treatment of the temporal variations in anthropogenic heat flux by the models can be sub-divided into three classes, those that neglect the anthropogenic heat flux completely, those that have fixed values in time, and those that have temporal variations (Table 4.2). For this stage of the comparison, no information about the anthropogenic heat flux (Q_F) was released to the participating groups, so the models used their own assumptions. This means that no advantage could be gained from any of the methods used to model Q_F due to knowledge of the observed variations in this flux.

The results for Q^* are similar for all three classes (Table 4.4 and Fig. 4.9 in supplementary material[10]), although the models that neglect Q_F generally have a smaller negative *MBE* than those that do include this term. Q_F influences Q^* through the outgoing longwave radiation link to surface temperature. These results suggest that the addition of Q_F increases the modelled radiative surface temperature, leading to a larger negative bias than is derived from the models that do not include this term, due to the increased outgoing longwave radiation. The models that include a temporal variation in Q_F have the smallest *RMSE* throughout the year, apart from the winter (June/July) period, with little seasonal variation (Table 4.4 and Fig. 4.9 in supplementary material). Of the other two classes that show a greater seasonal variability in the Q^* *RMSE*, those models with a fixed Q_F have the largest errors.

The models using a fixed Q_F also have a seasonal cycle in *MBE* for Q^*, with larger negative bias errors in the summer (Table 4.4 and Fig. 4.9 in supplementary material). It is possible that this source term is too large in the summer months leading to increased surface temperature and hence a larger negative bias in Q^*. However, a similar seasonal cycle in the biases for Q_H would be expected, but is not seen. Rather there is a minimum in the bias error at this time of year (Table 4.3, and Fig. 4.9 in supplementary material). This suggests that the treatment of Q_F is not responsible for the *MBE* in Q^* for these models, but it is likely to be other characteristics that these models share. Such a seasonal cycle is not present in the other two classes.

For Q_H, the models using a temporal variation in Q_F have a substantially larger *MBE* in summer compared to the other classes (Table 4.5 and Fig. 4.9 in supplementary material), but this bias reduces in winter to similar values in models that treat this with a fixed value. This is reversed for the *RMSE* where the models with a fixed Q_F have larger errors in the summer, but similar

[10] Best MJ and Grimmond CSB. Analysis of the Seasonal Cycle Within The First International Urban Land Surface Model Comparison – http://dx.doi.org/10.1007/s10546-012-9769-7.

errors in the winter compared to those using a temporal variation. At all times of the year the models that neglect Q_F have the smallest *RMSE* (Table 4.3, and Fig. 4.9 in supplementary material), however this is the only class that has a negative *MBE* in Q_H during the winter months. This is consistent with this class of models neglecting the additional available energy during the winter months, when Q_F has its greatest impact.

During the summer for Q_H, the models that do not include Q_F and those that have a fixed temporal behaviour to Q_F have similar daytime peak values for the diurnal fluxes (Fig. 4.10 in supplementary material), but positive biases compared to the observations. Those using a temporal variation in Q_F have higher peak values and hence larger positive biases than the other classes. This is also the case in winter, although the biases are smaller for all classes.

At night, models with a temporal variation in Q_F are in good agreement with observations at all times of the year (Fig. 4.10 in supplementary material), whereas the other classes have a negative bias. Interestingly the models that neglect Q_F have smaller negative biases during the night than those with a fixed value, despite the inclusion providing an additional source of energy that should be more significant at night. In autumn (Apr/May), the nocturnal biases are more substantial, except for the models that include a temporal variation that are still in good agreement with the observations (Fig. 4.10 in supplementary material). Hence having a seasonal variation to the anthropogenic heat fluxes may provide additional information that can reduce models errors at certain times of the year.

There is a similar peak difference in daytime mean fluxes for ΔQ_S with all classes of models, which implies the additional source of energy from Q_F has little impact during the day on ΔQ_S within the models. Nocturnally, the models with a fixed Q_F have the largest negative values of ΔQ_S, which is in better agreement with the observations in the summer, but has a negative bias during the winter (Fig. 4.10 in supplementary material). Hence models with a fixed Q_F have smaller nocturnal biases in ΔQ_S in the summer compared to the other classes but larger biases in the winter. However, in the second spring/summer at the end of the simulations, the biases for each of the categories of models are more similar to the winter, suggesting there may be some impact from model spin-up on this flux (not shown).

4.3.3 Categories that affect the radiation balance

For the various categories that influence the solar radiation at the surface, the simplest class (Table 4.2) consistently performs better than the other classes, with the lowest *RMSE* for all

fluxes over the whole of the seasonal cycle (Table 4.3, and Tables 4.4, 4.5 and 4.6 and, e.g., Fig. 4.11 in supplementary material). For the albedo and emissivity category, the single albedo models have a similar diurnal cycle to the observations for each flux over all periods, which is better than the other classes (not shown). Whilst the thermal properties for the materials of the various facets were provided to participants only the observed mean albedo and emissivity for the site were available (see Table II in Grimmond et al. 2011). This may explain why the models with a bulk albedo perform better than models that require albedo information for each facet (Grimmond et al. 2011).

The models with the simplest characteristics in the facet and orientation category are exactly the same as the models with the simplest characteristics in the albedo and emissivity category. Hence specifying the observed mean albedo and emissivity for the site also affects the results for the facet and orientation classes, with the models that have the simplest characteristics performing the best. These models have only a small seasonal cycle in their *MBE* for ΔQ_S (Table 4.6 and Fig. 4.11 in supplementary material), but whilst they have the smallest negative *MBE* in this flux during summer at the start of the simulation, there is an offset compared to the observations that persists during the seasonal cycle. This is not seen for the other model classes in either of the categories, which have an improvement in *MBE* during the winter.

Many of the models with the simplest radiation characteristics have the simplest approach to reflections. Although bulk albedo models are not the majority in the class they influence the median and hence the class results. Provision of the bulk albedo, which is a direct input for the simplest radiation categories, may be of added advantage to these model types. However, obtaining a bulk albedo for an urban area is probably easier using global satellite data than obtaining specific material albedos within the city requiring detailed local knowledge.

The more complex classes within each of these categories affect the results. For the facets and orientation category, the models with orientation but no intersections have a larger *MBE* for Q_H than the other model classes, but this is not seen in the *RMSE* (Table 4.3, and Fig. 4.11 in supplementary material). This larger positive bias is not due to greater peak fluxes during the diurnal cycle, but from a phase shift in this flux that leads to a delay in the peak and subsequent reduction of Q_H in the afternoon.

Also with the facet and orientation category, the models that include intersections have a diurnal amplitude of ΔQ_S that is substantially smaller than that observed (not shown). Whilst the amplitude of the diurnal cycle of ΔQ_S is also smaller than observed for some of the other classes, those models with intersections are the only cohort with both a positive nocturnal bias

and a negative daytime bias, which is seen consistently throughout the seasonal cycle.

The results when categorizing the models with respect to the number of reflections for solar radiation, show that both the multiple and infinite reflection classes overestimate the peak daytime Q_H in winter (Fig. 4.12 in supplementary material). For the former, an underestimation of Q_E compensates for Q_H, giving a ΔQ_S that is similar to the single reflection models. However, in the latter case it leads to a greater negative bias in the peak ΔQ_S.

Finally for the albedo and emissivity category, the cohort that uses two albedos substantially overestimate Q_H compared to the observations and the other approaches (Table 4.3). In winter, when the other classes have a near zero bias, the two albedo models still have a *MBE* > 30 W m^{-2} (Table 4.3) whilst in summer the *RMSE* > 130 W m^{-2}. This is larger than the mean observed flux during the summer months which is less than 100 W m^{-2} (Table 4.3). The diurnal pattern has an overestimate of the peak daytime Q_H throughout the year, which is as much as 100% in the summer (not shown). The larger values of Q_H result from an overestimation of peak Q^* during the day. The apparent inconsistency of a small negative bias in Q^* is explained by the faster reduction in Q^* in the afternoon, leading to a negative bias during this period that offsets the positive bias at the peak flux (not shown).

4.3.4 Method to calculate Net Storage Heat Flux

Although there is a clear seasonal cycle in the net storage heat flux in both the observations and the models, it is not clear that the method used for calculating this net storage heat flux would itself lead to seasonal variations in the performance of the models. However, the seasonal cycle of the fluxes and the errors for this category show some distinct differences in the seasonal behaviour of the different classes of models (e.g., Table 4.3, and Table 4.6 and Fig. 4.13 in supplementary material).

The models that use Q^* to calculate ΔQ_S have a different behaviour to the other two classes with little seasonal cycle in the *MBE* for Q_H (Table 4.3, and Fig. 4.13 in supplementary material). This leads to a smaller *MBE* in summer and larger bias in winter compared to the other model classes. This cohort has the smallest *RMSE* through the year, although the other classes are similar in winter (Table 4.3, and Fig. 4.13 in supplementary material). For ΔQ_S itself, this group has a close to zero *MBE* in summer, when the others are at their maximum. However, in winter all classes have their reverse (Table 4.6 and Fig. 4.13 in supplementary material). As the actual observed mean flux is smaller in winter, we might expect a *MBE* minimum, as occurs for the

models that calculate ΔQ_S as a residual and those that use the heat conduction equation.

The Q^* based models have a similar diurnal pattern as observations in the summer for ΔQ_S, but in winter ΔQ_S has a negative bias throughout the diurnal cycle (e.g., Fig. 4.4). Results for each 60-day period are shown in Fig. 4.14 in the supplementary material. This contrasts with the other classes that are in good agreement with the observations during the night in winter. In addition to having a better diurnal amplitude during summer, the Q^* based cohort is the only group with the correct phase for ΔQ_S (Fig. 4.4).

As the Q^* based methods derive their parametrization from correlating component facet ΔQ_S with observations of Q^*, and most observational studies have been conducted in summer months, this probably means there is a bias towards this time of the year. Hence these models perform substantially better in summer than in winter. This indicates that users need to ensure the parameters used are appropriate seasonally; for example the Anandakumar (1999) type results may be very important to include, in order to prevent extrapolation beyond the information used in its derivation.

4.3.5 Model Complexity

For the complexity category, the results show that the simple models generally have the smallest *RMSE* throughout the year for all of the fluxes, and the smallest *MBE* for both Q_H and Q_E compared to the other two complexity classes (Table 4.3, and Tables 4.4, 4.5 and 4.6 and Fig. 4.15 in supplementary material). For both *MBE* and *RMSE*, the complex models have the largest errors in Q_H during summer, but similar errors to the medium complexity models during winter. The two more complex classes have similar *MBE* and *RMSE* for Q_E throughout the year with larger errors in summer than winter, whilst the most complex models have the largest *RMSE* for ΔQ_S at all times.

As with the majority of other categorisations, the mean flux is close to zero for ΔQ_S throughout the year for all three classes (e.g., Fig. 4.2). This behaviour is not seen in the observations where there is a positive flux in summer. It is difficult to fully analyse this result because of missing observations, although it does suggest that forcing energy balance closure through ΔQ_S over relatively short periods may not be the observed behaviour of the urban system. However, there will be a bias in the observations to dry periods so the rapid loss of heat during rainfall is not observed, helping to contribute to a positive bias in the observations.

For all of the fluxes, except possibly Q_E, the three classes are able to capture the general seasonal variations in the diurnal cycle (e.g., Fig. 4.5. Results for each 60-day period are shown in Fig. 4.16 in the supplementary material). As there are still few datasets that cover all periods of the year, this is a good demonstration that the models, in general, are able to capture the main physical processes for urban environments. For Q_H, the differences in diurnal cycle across the complexity classes have a similar pattern through the year (Fig. 4.5). All overestimate the peak daytime fluxes, except for the simple models in winter months which agree well with the observations. The biases in peak daytime Q_H increase with complexity (simple to medium to most complex) but in winter, the biases in complex models are similar to the medium complexity models (Fig. 4.5).

The amplitude of ΔQ_S varies between the differing complexity classes (Fig. 4.5), but is consistent throughout the year with the complex class having the smallest amplitude in diurnal cycle throughout the seasons. A larger amplitude is simulated by medium complexity models which is in better agreement with the observations at all times of the year.

To assess the affect of vegetation on the results for the overall complexity of the models, the results were re-calculated with the models that do not include a representation of vegetation removed from the analysis. There is no class in the complexity category that does not have at least one model which neglects vegetation, but the distribution of these models within the classes could have a detrimental impact on one class more than the others. However, the resulting analysis shows that, with the exception of the *RMSE* for Q^* for which the performance of the classes is reversed, the remaining overall results are unaltered with the simple models performing the best, followed by the medium complexity models.

4.3.6 Outlying Models

Whilst comparing the median data points for the *MBE* and *RMSE* allows an overall assessment of the different representation of the physical processes in each of the categories, these analyses deliberately exclude outlying models. Now we assess the characteristics of all models, including the outliers, by comparing of each point to determine the percentage of model data that are within a specified tolerance of the observed values. An example is shown in Fig. 4.6, where the y-axis is the percentage of model results within the given observation tolerance of the x-axis.

Two methods are used to provide the tolerance values: (1) percentage of the observed flux being compared; and (2) percentage of the observed Q^*. The latter provides a basis to calculate the

observational errors for each of the directly measured fluxes (Hollinger and Richardson 2005), with ΔQ_S having the sum of the observational errors of the other fluxes as it is often calculated as the residual of the surface energy balance. A tolerance value of 20% of the flux has been used for both the observed flux method and the observed net radiation method to assess a ranking for the models over each flux for the whole seasonal cycle (shown by the vertical dotted line in Fig. 4.6).

In general, the models have a higher percentage of points that are closer to the observed values in summer than in winter (Fig. 4.6). However, this could be a result of the larger tolerance values that are obtained because the observed fluxes are larger in summer. Comparing the tolerance values derived from the observed flux method or from the observed Q^* method shows that there is little difference in the resulting model behaviour (not shown). There is some re-ordering of the models, but they are still grouped around the same position in a ranked list. The exception to this is for Q_E where there is a substantially different curve shape from the models with a tolerance based on the observed Q_E, compared to the shape with a tolerance based upon the observed Q^* (not shown). This change in shape is determined by the size of the tolerance, since the observed Q_E is substantially smaller than the observed Q^* during the daytime. However, despite the change in the shape of the curve, there is no significant change in the ranked ordering of the models.

Once the models are ranked, it is possible to assess what similar characteristics the best (and worst) performing model share. In general, however, it is found that there is no clear distinction, with a particular characteristic of top ranked models being shared with the models near the bottom. This is especially true for Q^* where there are no obvious classifications in the physics of the models that describe the rankings, nor the changes to the rankings over the seasonal cycle (not shown). Whilst the worst performing model has a different shape curve to the other models (Fig. 4.6), it shares the same classifications with other models. This shows that the physical categories considered in the analysis can not explain all of a model's behaviour. Other elements, such as the way in which the options have been implemented, can affect the model performance.

For the daytime Q_H, the shape of the curves fall into three types (Fig. 4.6): the best performing have a concave shape, the worst performing a convex shape, whilst the remaining models have an inflection in the curve. For all but two of the models that do not include vegetation, these models make up the worst performing models with the convex shape. The remaining two models that neglect vegetation perform better than some models that do include a representation of vegetation. However, these two models include an alternative moisture source and hence still have some evaporation that prevents the excessive sensible heat flux during the daytime.

For the other categories there is a mixed signal in terms of the classifications, but some general conclusions can be observed. The models with a convex shape tend to neglect a representation of vegetation and have multiple reflections, although not all models with these attributes have such a convex shape. Also, there is a tendency for models with an inflection to have infinite reflections. In general for summer, the classification of reflections has some affect in the model rankings (not shown). On average the best models have a single reflection, the worst performing models have multiple reflections, whilst the models with infinite reflections are in the middle of the rankings. In winter the signal is more mixed between the multiple and infinite reflections. However, models with single reflections still tend to be amongst the better performing models. One characteristic that the best performing models share on the whole is that they do not include an anthropogenic heat flux (not shown). This holds for all of the seasonal periods including winter.

For Q_E, again the worst performing models generally neglect vegetation (not shown). When comparing the model results using a tolerance based upon the daytime observed Q_E, all except two of the models that neglect vegetation do not have any points within 100% of the observed values. The two remaining models without vegetation are the ones with an alternative moisture source for evaporation, and one of these performs better than some of the models that have a representation for vegetation throughout the seasonal cycle. This demonstrates that having a representation of vegetation is necessary, but not sufficient, to model Q_E well.

Models that calculate ΔQ_S from Q^* are near the top of the rankings for the daytime net storage heat flux during summer, but fall to the bottom half of the rankings during winter (not shown). However, when considering the results for the nighttime ΔQ_S, these models are not near the top of the rankings at any time during the year, even for summer. In addition, models with more than two albedos tend to be grouped near the bottom of the model rankings for the daytime ΔQ_S.

To assess the impact of observational errors on the results showing that the simple models perform the best, the models have been classified into the simple, medium complexity and complex classes in Fig. 4.6 (note that results for each individual model are identified in Fig. 4.17 in the supplementary material). The ranking shown is based upon the number of model points within 20% of the observed flux, which are the values that Coutts et al. (2007b) suggest the sensible and latent heat fluxes could be under-estimated. Whilst no class obviously performs better than another, these results show that the spread of models in the simple class tend to be amongst the better performing models which explains why the median model is better in this class. For the complex class there are fewer models that are better performing and hence the

median model has the relatively worse performance.

4.3.7 Phase Errors

The impact of a phase error in the modelled fluxes can be assessed by shifting the model results temporally and re-comparing against the observed fluxes. Here the model results are shifted one (30-min) and two (60-min) timesteps forward, and one (-30-min) and two (-60-min) timesteps backwards. An example of this analysis is shown in Fig. 4.7.

For Q^* most models have their best performance without a time shift (e.g., two of the models in Fig. 4.7. Results for all of the models are shown in Fig. 4.18 in the supplementary material), but there are a few (seven) models that do have some benefit from a time shift. This behaviour holds during all of the seasonal cycle except for one model during summer (February/March) which then has the best performance without a phase shift. These models either have a consistent phase lag or lead (three models), or have an inflection without a time shift (four models). An example of each is shown in Fig. 4.7. These latter models show improved results from a time shift at low tolerance values, but for the larger tolerance values have the best results with no time shift.

The situation is quite different for Q_H, where most models perform better with a time shift. The direction of the time shift giving better results for each model is generally the same over the seasonal cycle, but there is no consistency in the direction of the time shift between models; i.e., some perform better when leading and some perform better when lagged. This is also the case for ΔQ_S with few models performing best without a time shift and no consistency in the direction of the time shift. In addition, there is no consistency for the models between the phase shift for Q_H and ΔQ_S (not shown).

For Q_E, all of the curves for a given model are very similar. Hence there is no real difference between having a time shift with a lead or a lag, or having no time shift at all (not shown). This probably results from the small daytime latent heat flux values, giving a flatter diurnal shape to the time series.

Neglecting Q_E (as all of the time shifted curves are close together), there are only six models that consistently give the best performance with no phase shift over all seasons for all of the fluxes. Of these models, all but one model has a tile representation for vegetation, with the remaining model neglecting vegetation completely. The one physical characteristic that is shared by all of these models is that they all neglect the anthropogenic heat flux (not shown).

One of the models uses the residual method to calculate ΔQ_S, whereas all of the others use the diffusion equation. In terms of overall complexity, all but one of the models is classified as simple, whilst the remaining model is classified as medium complexity.

Whilst this analysis shows which fluxes within each of the models have a time shift compared to the observations, it is not possible for any of the models to adopt different time shifts for each of their fluxes, as they are intrinsically linked through the basic surface energy balance equation. Hence an overall analysis of the affect of a given time shift on all of the fluxes is required. To do this, a combined sum is used to identify the number of model points that are within the error estimate of the observed flux. The observational error estimates for the fluxes are taken as a fraction of the observed net radiation, with these fractions taken from Hollinger and Richardson (2005), i.e., 5% for Q^*, 10% for Q_H and 8% for Q_E. As the net storage heat flux is determined as the residual of the energy balance within the observational dataset, this flux is omitted from the analysis.

An example of the resultant behaviour for the models is shown in Fig. 4.8 (results for all models are shown in Fig. 4.19 in the supplementary material) with the x-axis indicative of the combined observational error. A value of one gives the summed observational errors, whilst a value of two is twice the observational errors, etc. The results show that the time shift for the combined measure is dominated by the signal for the net radiation, with most models having their best performance with no time shift (Fig. 4.7, 4.8). In addition, the seven models identified as benefiting from a time shift for their net radiation are the same as the models that have some benefit from a time shift within the overall measure. This is an indication that the net radiation is the best represented flux within the models in general, and demonstrates that on the whole, the models do not have a dominating phase error.

4.4. Conclusions

Given the limited availability of urban observations that are for a sufficiently long period to study the seasonal cycle, this analysis using data from the first international urban model comparison provides unique insight into the seasonal behaviour of the various types of urban land-surface model. By grouping the models into a number of different categories based upon their representation of various physical processes, it is possible to assess the processes that are the most critical for representing an urban environment.

We conclude that, in general, urban models can represent the seasonal cycle in the surface

fluxes. The models tend to have larger errors in summer than in winter, as is expected given the larger magnitude of the insolation in summer, and of the fluxes generally. Normalizing these errors by the mean observed flux gives a similar response throughout the year, except when the magnitude of the observed flux is small during the winter months. Most models have their best performance without a time shift compared to the observations, implying that they do not have an overall phase error. Q^* has the smallest errors at all times of the year, and in general the models have a negative bias through the seasonal cycle for this flux. The models also have a negative bias in Q_E and ΔQ_S, whereas Q_H generally has a positive bias throughout the year.

On a seasonal basis the physical process most critical to represent in an urban area is the evaporation from vegetation. Models that do not include this perform more poorly than those that do include vegetation. When comparing the number of model data points within a given tolerance of the observed daytime Q_E, all except two of the models that neglect vegetation do not have any data within 100% of the observed values, and these two models include an alternative moisture source for evaporation. However, the method of inclusion is not as important; there is little difference between the tile scheme and integrated approaches, except that only models including the tile approach give better results without a phase shift in their fluxes. The errors in Q_E resulting from models that neglect vegetation lead to compensating errors in Q_H. For summer, the *RMSE* for both Q_H and Q_E from these models is larger than the observed mean fluxes. Hence these models show poor performance for both turbulent fluxes, which provide the main driver for the evolution of the atmospheric boundary layer. Despite these overall results one of the models that neglects vegetation, but has an alternative moisture source for evaporation, still appears higher in the rankings for Q_E than several models that do include a representation of vegetation, at all times of the year. Hence a representation of vegetation within urban areas is necessary, but not sufficient, for modelling Q_E realistically.

Typically, simple models perform better than more complex models for all of the fluxes over the seasonal cycle, within each of the other categories considered. In particular, for Q^* where we might expect to see more benefit from the treatment of radiation at the surface, we find that the models that represent the urban area as a bulk form, or single facet, perform better that those with many facets, with or without orientation, or intersections. Models that use a bulk albedo perform better than those that use separate albedos for multiple surfaces. Those that simulate a single reflection have a smaller *RMSE* than those with multiple or infinite reflections, although they have a larger negative *MBE*. As only the observed bulk albedo was available to the modelling groups, it is possible that only the simple models in each of these categories were able to take full advantage of this information. However, it is not unreasonable to assume that it would be easier in general to deduce a bulk albedo for a city from available satellite data, than

to derive the specific albedo of the building materials.

Across all of the categories considered, models from all classes tend to simulate a mean ΔQ_S close to zero throughout the seasonal cycle, except for those that use Q^* to calculate ΔQ_S. This small flux is not seen in the observations and could imply that the models that force energy balance closure through ΔQ_S are doing it at too short timescales compared to those that actually occur in the urban environment.

The models that use Q^* to calculate ΔQ_S have the best performance for this term in the surface energy balance compared to the other models, during the daytime in summer. However, these models have a seasonal cycle in the *MBE* that is opposite to the other types of models, with smaller negative *MBE* in summer and larger *MBE* in winter. This could result from the coefficients used to calculate ΔQ_S being biased from summer based observations. Hence these models would benefit from longer term multi-annual observations within urban areas from which new empirical relationships could be determined.

In general, the models that neglect the anthropogenic heat source do at least as well as the models that do include this term, despite missing this additional energy source for the surface energy balance. Whilst these results seem to be counter intuitive and contradict other studies, it is not possible to explain these results based upon other obvious categories that these models have in common. Further analysis of these results are beyond the scope of this paper, but will be the study of future work. Models that incorporate a temporal variation in Q_F show some increased skill at nighttime during the winter, and these are the only models with Q_H similar to the observations during this period. However, during the daytime these models consistently overestimate Q_H at all times of the year.

When the models' complexity is the basis for classification, the results show that the simple models give the best simulation for all fluxes. These models have the smallest *RMSE* for each flux and the smallest *MBE* for Q_H and Q_E, and form the majority of models that do not benefit from a phase shift in any of their fluxes. However, simple models do have a larger bias in Q^*, although this is the most accurately represented flux within the surface energy balance. Again it is not possible to identify classes of models from other categories that could result in this behaviour. There are simple models in each class for all of the categories. It is worth noting that the complex class are defined as the models that have no simple characteristics over all of the physical categories, where as the medium complexity models have only one or two simple characteristics. Hence it is possible that interaction between the complex characteristics lead to a degradation of skill within these models. The models in the simple class have the largest

number of simple characteristics and hence interactions between their physical schemes are easier to understand.

In general, apart from the vegetation category, the simple models have the best representation of the diurnal cycle for Q_E, including the models that neglect Q_F when classified by the representation of this term. The restricted moisture fluxes from the more complex models result in a compensating over estimation of Q_H, again leading to greater errors. Thus representing evaporative fluxes from vegetation has a larger affect on the accuracy of the surface fluxes than adding more complex details to the representation of the urban surface components. However, it is likely that this could be explained by the representation used for vegetation rather than the other physical properties. It is possible that the simple models include a representation of vegetation, whereas the more complex models have concentrated on the complexity of the urban built morphology, design, construction materials and anthropogenic heat sources.

Given a set of classifications that the best models share, there are also other models with the same classifications that perform less well (a list of the cross classifications are given in Grimmond et al. 2010). Hence despite the general conclusions about various physical classes that can be made for the models, other elements, such as the implementation of these options, can also affect the overall model performance.

Whilst the comparison considered model results across two sites, only one of these was used for the results in Phase 2/Stage 4 of the comparison, which enabled a clean analysis of the physics within the models, as no evaluation data were available to the modelling groups. Hence it is not known if these conclusions are robust for other sites with differing density of buildings and complexity in the urban morphology. It is therefore recommended that additional observational sites should be used to corroborate the findings of this study. However, it is noted that it is not easy to obtain good quality data within the urban environment, given the limitations on measurement locations and representative spatial extents.

Table 4.1: The number of model versions used in stage 4 of the first international urban model comparison (Grimmond et al., 2010, 2011)

Code	Model name	References	Versions	Groups
BEP02	Building effect parametrization	Martilli et al. (2002)	1	1
BEP_BEM08	BEP coupled with building energy model	Martilli et al. (2002); Salamanca et al. (2009, 2010) ; Salamanca and Martilli (2010)	1	1
CLMU	Community land model – urban	Oleson et al. (2008a, 2008b)	1	1
IISUCM	Institute of Industrial Science urban canopy model	Kawamoto and Ooka (2006, 2009a, 2009b)	1	1
JULES	Joint UK land environment simulator	Essery et al. (2003); Best (2005); Best et al. (2006); Best et al. (2011)	4	2
LUMPS	Local-scale urban meteorological parametrization scheme	Grimmond and Oke (2002); Offerle et al. (2003); Loridan et al. (2011)	2	1
NKUA	University of Athens model	Dandou et al. (2005)	1	1
MORUSES	Met Office Reading urban surface exchange scheme	Harman et al. (2004a, 2004b); Porson et al. (2010)	3	1
MUCM	Multi-layer urban canopy model	Kondo and Liu (1998); Kondo et al. (2005)	1	1
NJU-UCM-S	Nanjing University urban canopy model single layer	Masson (2000); Kusaka et al. (2001)	1	1
NSLUCM/ NSLUCMK/ NSLUCM-WRF	Noah land surface model/single-layer urban canopy model	Kusaka et al. (2001); Chen et al. (2004); Loridan et al. (2010)	3	3
SM2U	Soil model for sub-mesoscale (urbanized)	Duport and Mestayer (2006); Dupont et al. (2006)	1	1
SNUUCM	Seoul National University urban canopy model	Ryu et al. (2011)	1	1
SRUM2/ SRUM4	Single column Reading urban model tile version	Harman and Belcher (2006)	4	1
SUEB	Slab urban energy balance model	Fortuniak (2003); Fortuniak et al. (2004, 2005)	1	1
SUMM	Simple urban energy balance model for mesoscale simulation	Kanda et al. (2005a, 2005b); Kawai et al. (2007, 2009)	1	1
TEB	Town energy balance	Masson (2000); Masson et al. (2002); Lemonsu et al. (2004); Pigeon et al. (2008)	1	1
TEB-ml	Town energy balance with multi-layer option	Hamdi and Masson (2008)	1	1
TUF2D	Temperatures of urban facets 2D	Krayenhoff and Voogt (2007)	1	1
TUF3D	Temperatures of urban facets 3D	Krayenhoff and Voogt (2007)	1	1
VUCM	Vegetated urban canopy model	Lee and Park (2008)	1	1

Note: These are assigned anonymous numerical identifiers in the analysis

Table 4.2: Number of models in each class (simple 1 to medium 2 to complex 3 or 4) by model category (see Fig. 4.1 in Grimmond et al. 2011) considered for the seasonal cycle analysis. See text for further explanation

Category	Class		
		Name	No. of models
Vegetation	1	None	8
	2	Separate tile	19
	3	Integrated	5
Temporal variation of the anthropogenic heat flux	1	None	22
	2	Fixed	3
	3	Variable	7
Facets & orientation	1	Bulk	5
	2	Roof, walls, road, without orientation	17
	3	Roof, walls, road with orientation, no intersection	6
	4	Roof, walls, road with orientation, with intersection	4
Reflections	1	Single	11
	2	Multiple	13
	3	Infinite	8
Albedo, emissivity	1	Bulk	5
	2	Two facets	4
	3	Three facets	23
Calculation method for the net storage heat flux	1	Net radiation	3
	2	Residual	6
	3	Conduction	23
Complexity	1	Simple	14
	2	Medium	12
	3	Complex	6

Table 4.3: Seasonal cycle (60-day averages) of the mean bias error (MBE) and root mean square error (RMSE) for the sensible heat flux (Q_H) within each class in a number of categories (Table 4.2), defined by the model representation of the physics

		RMSE							*MBE*						Q_H
Mean Obs		90	93	59	8	-4	14	39							
		O/N	D/J	F/M	A/M	J/J	A/S	O/N	O/N	D/J	F/M	A/M	J/J	A/S	O/N
Vegetation	1	133	134	106	45	35	57	96	68	63	45	24	22	34	52
	2	66	68	61	31	24	32	50	21	18	11	-3	-7	-1	10
	3	57	61	48	30	23	32	45	18	23	11	2	-8	-1	8
Temp. Var. Q_F	1	61	67	61	30	24	32	45	18	20	15	-2	-6	0	10
	2	94	99	88	39	35	41	72	21	11	8	8	15	18	13
	3	85	88	79	36	34	41	60	40	42	34	10	10	17	22
Facets & orientation	1	57	56	47	25	22	26	38	10	11	0	-7	-2	-1	-1
	2	73	72	66	32	28	33	52	32	26	17	0	-7	3	14
	3	77	79	65	33	28	41	53	39	37	27	8	12	17	22
	4	95	99	86	47	37	57	78	21	23	-1	-10	-5	-5	3
Reflections	1	57	56	48	28	22	26	38	10	11	6	-6	-2	-1	7
	2	95	99	86	44	34	55	78	39	37	27	5	10	17	8
	3	77	79	65	33	28	33	53	36	26	17	0	-5	3	14
Albedo, Emissivity	1	57	56	47	25	22	26	38	10	11	0	-7	-2	-1	-1
	2	137	134	106	50	42	63	100	80	74	56	33	33	45	63
	3	74	72	65	33	28	33	52	32	26	17	2	-5	3	13
ΔQ_S calculation	1	61	61	48	30	26	33	46	10	11	8	12	13	15	20
	2	68	71	65	31	25	33	51	25	26	19	2	-5	3	13
	3	73	72	64	35	28	33	52	32	26	16	-2	-6	0	10
Complexity	1	57	58	49	28	22	26	39	13	12	6	-3	-5	-1	8
	2	77	79	65	39	32	41	60	36	28	19	8	12	17	22
	3	109	109	88	44	35	55	78	44	42	34	12	10	17	20

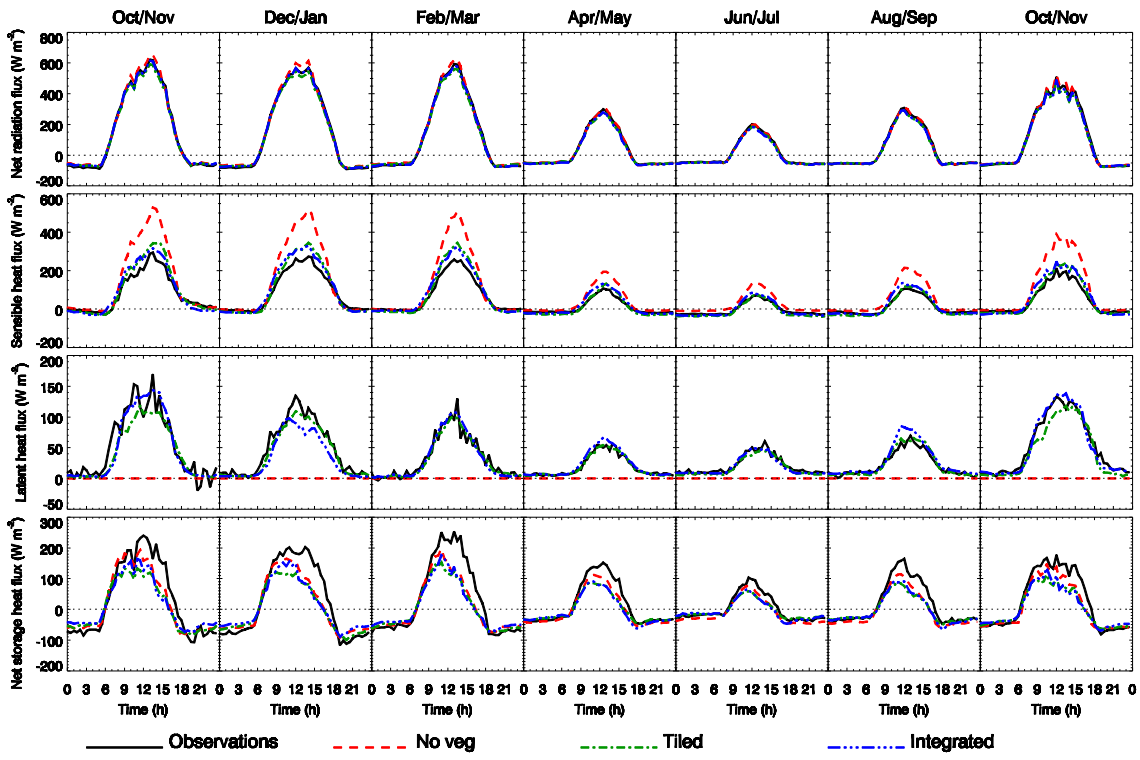

Figure 4.1: Median of the average diurnal cycle for each 60-day period throughout the seasonal cycle, from the models classified by their representation of vegetation. Note the scales are different for each flux.

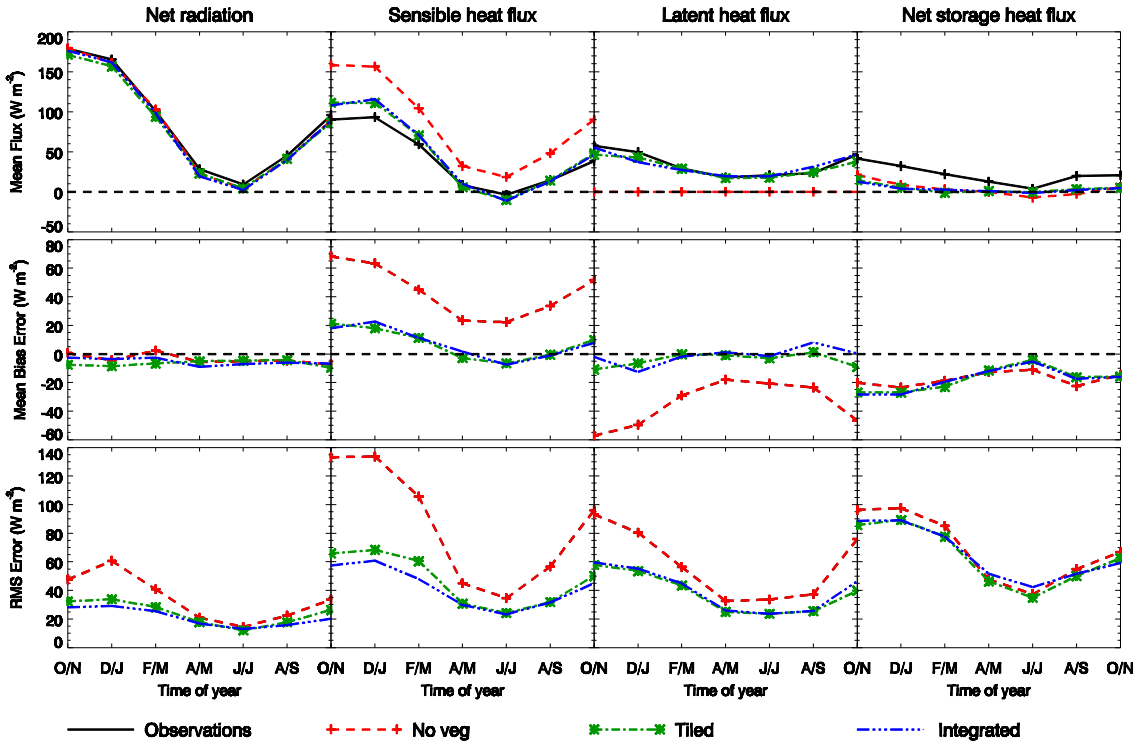

Figure 4.2: Median of the mean modelled flux, mean bias error, and root-mean-square error for the surface fluxes determined for two month periods, for the models classified by their representation of vegetation. Note the scales are different for each flux.

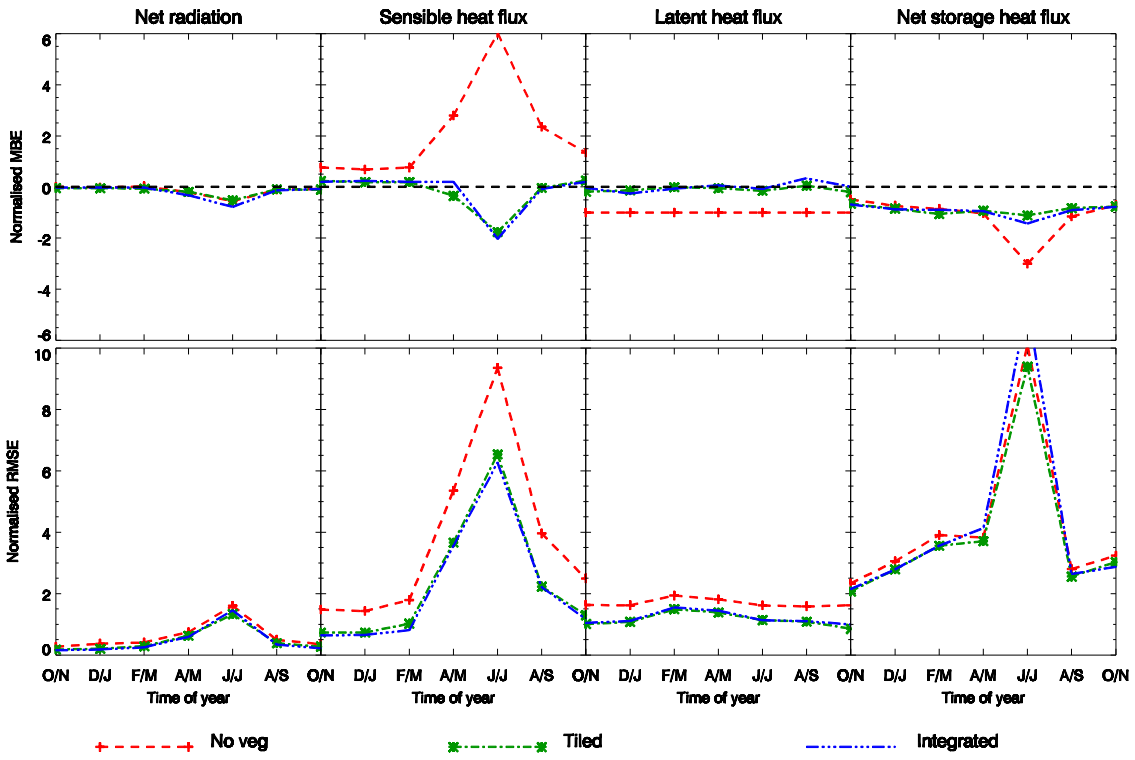

Figure 4.3: Normalized median mean bias error (MBE) and root-mean-square error (RMSE) for the surface fluxes classified by their representation of vegetation in the model. The errors are normalized by the absolute value of the mean observed fluxes for the appropriate period in the seasonal cycle. Note the scales are different in the two rows.

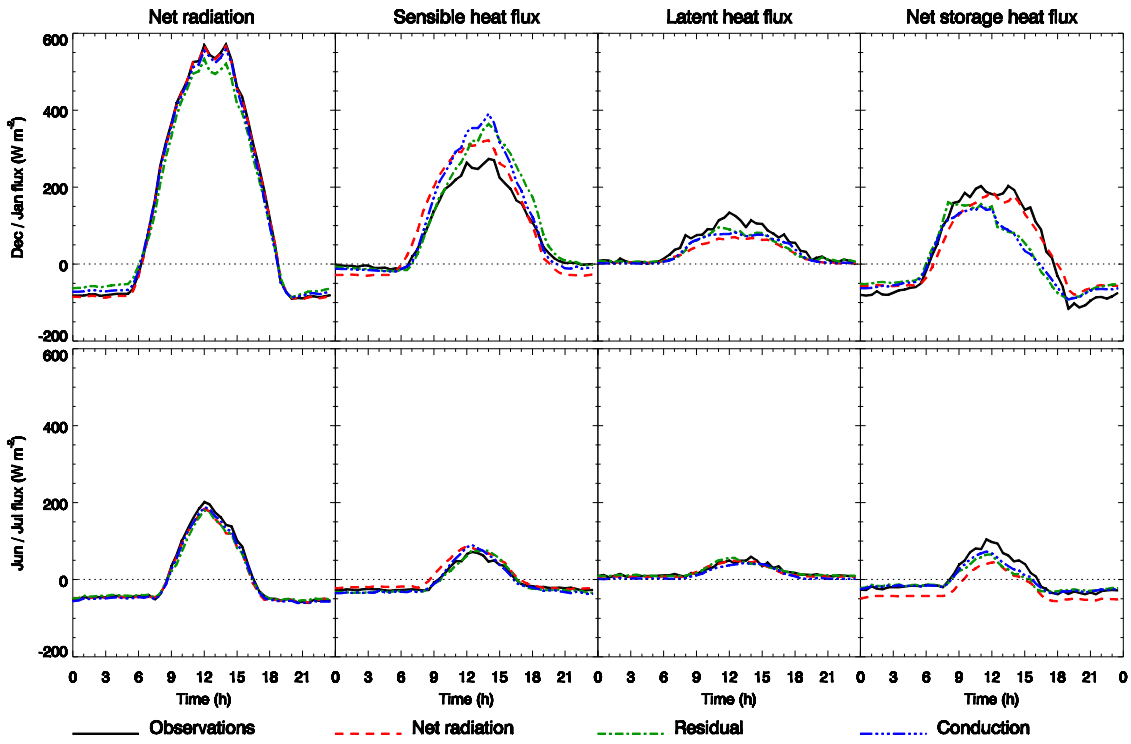

Figure 4.4: As Fig. 4.1 for an example summer and winter 60-day period, but with the models classified by the method for calculating the net storage heat flux.

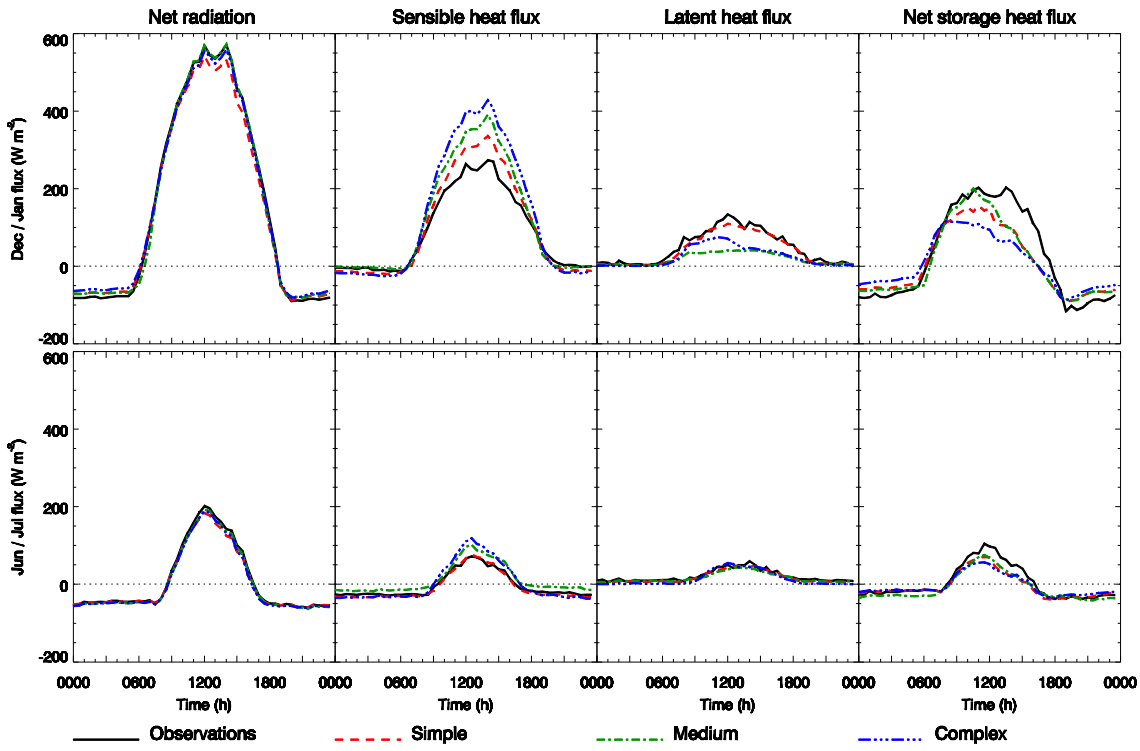

Figure 4.5: As Fig. 4.1 for an example summer and winter 60-day period, but with models classified by their complexity.

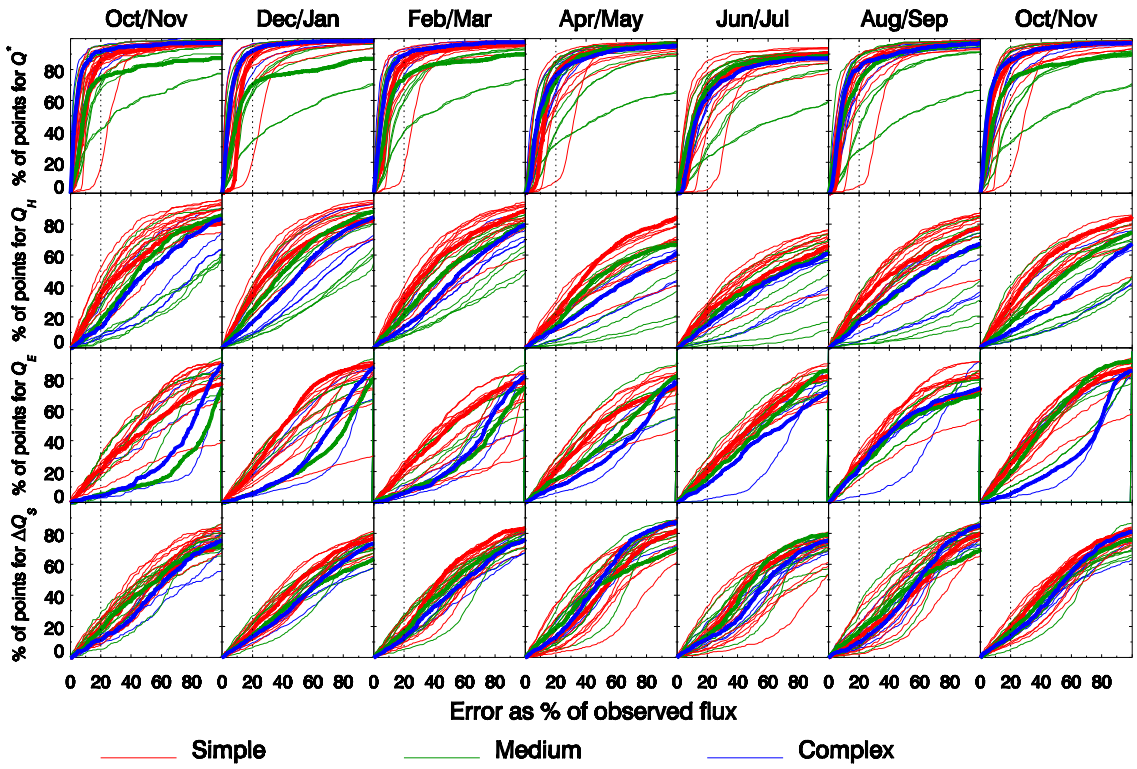

Figure 4.6: Percentage of modelled data within a percentage of the observed flux for each individual model, classified by their complexity, for each flux and 60-day period. An observational error estimate of 20% (the value for the possible under estimation in the sensible and latent heat fluxes as suggested by Coutts et al. 2007b) is indicated by the vertical dotted line. Thick lines indicate the median model when ranked at the 20% tolerance.

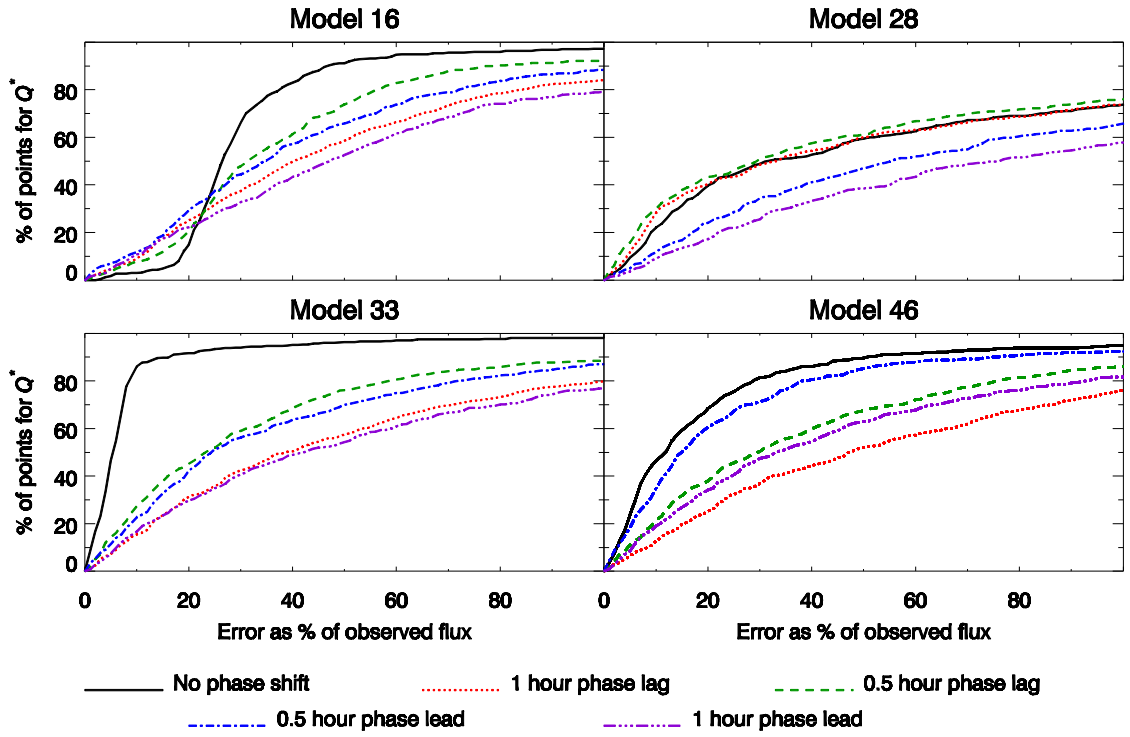

Figure 4.7: As for Fig. 4.6 for some example models, but with the individual results shown for net radiation for the February-March period. The modelled results have been shifted relative to those returned by the modelling group: 30 min and 60 min earlier and later.

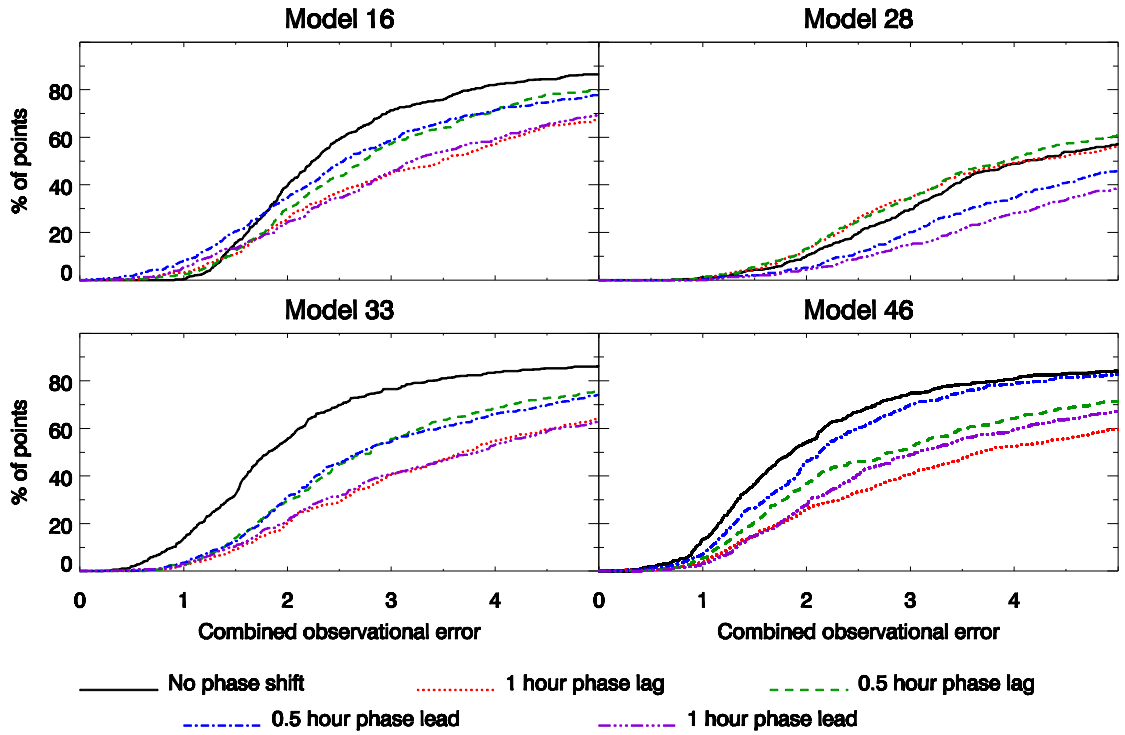

Figure 4.8: As for Fig. 4.7, over the February-March 60-day period, but for the model data combined using the observational error estimates given by Hollinger and Richardson (2005) as a fraction of the observed net radiation, for the net all-wave radiation (5%), the sensible heat flux (10%) and the latent heat flux (8%). The net storage heat flux is omitted as it is determined from the residual of the energy balance within the observational dataset.

Supplementary material for:

Best, M.J., C.S.B. Grimmond (2013), Analysis of the seasonal cycle within the first international urban land surface model comparison, *Boundary-Layer Meteorol.*, 146, 421-446, doi: 10.1007/s10546-012-9769-7.

Table 4.4: Seasonal cycle (60-day averages) of the mean bias error (MBE) and root mean square error (RMSE) for the net all wave radiation (Q^*) within each class in a number of categories (Table 4.2), defined by the model representation of the physics.

		Mean Obs	179	165	101	28	9	46	96	*MBE*						*Q**
		RMSE	O/N	D/J	F/M	A/M	J/J	A/S	O/N	O/N	D/J	F/M	A/M	J/J	A/S	O/N
Vegetation	1		48	61	41	21	14	23	34	-1	-4	2	-6	-5	-5	-7
	2		32	34	28	18	12	18	26	-8	-9	-7	-5	-5	-4	-10
	3		28	29	25	17	13	16	20	-3	-4	-3	-9	-7	-6	-7
Temp. Var. Q_F	1		32	34	29	17	12	18	26	-4	-4	-1	-3	-4	-4	-7
	2		33	40	36	28	21	23	32	-1	-21	-2	-5	-5	-2	-9
	3		19	20	18	13	13	13	15	-8	-9	-4	-8	-8	-1	-10
Facets & orientation	1		20	22	18	17	11	16	18	-5	-4	-6	-6	-7	-10	-10
	2		32	34	28	17	12	18	26	-1	-5	-1	-3	-4	-2	-6
	3		37	37	30	21	14	23	30	-8	-9	-4	-8	-8	-10	-10
	4		51	50	38	25	20	30	41	-1	-0	-3	-8	-7	-11	-9
Reflections	1		25	27	21	17	11	16	21	-9	-11	-7	-8	-7	-9	-11
	2		40	39	32	19	14	19	32	-3	-5	-2	-6	-5	-5	-7
	3		33	35	28	20	16	23	29	-6	-5	-4	-3	-4	-2	-1
Albedo, Emissivity	1		20	22	18	17	11	16	18	-5	-4	-6	-6	-7	-10	-10
	2		50	64	42	17	14	19	36	-8	-9	-3	-3	-1	-4	-8
	3		32	35	29	19	13	18	29	-2	-4	-2	-7	-6	-6	-7
ΔQ_S calculation	1		32	35	31	28	21	19	29	-5	-4	-6	-5	-4	-3	-9
	2		47	50	42	34	30	38	45	-5	-5	-2	-7	-7	-6	-8
	3		28	29	25	15	12	17	22	-4	-5	-3	-6	-5	-5	-8
Complexity	1		32	34	29	18	12	18	26	-5	-8	-7	-6	-6	-6	-10
	2		48	50	41	18	14	19	36	-2	-0	8	-1	-4	-2	0
	3		33	35	28	20	14	23	29	-1	-1	0	-7	-8	-10	-7

Table 4.5: As Table 4.4, but for the latent heat flux (Q_E)

		RMSE O/N	RMSE D/J	RMSE F/M	RMSE A/M	RMSE J/J	RMSE A/S	RMSE O/N	MBE O/N	MBE D/J	MBE F/M	MBE A/M	MBE J/J	MBE A/S	Q_E O/N
Mean Obs		57	50	29	18	21	24	47							
Vegetation	1	93	80	56	33	34	37	76	-57	-50	-29	-18	-21	-24	-47
	2	58	54	44	25	24	26	40	-11	-7	0	-1	-3	1	-9
	3	60	55	45	26	24	26	47	-2	-13	-2	1	-2	8	0
Temp. Var. Q_F	1	59	54	45	26	24	25	40	-5	-8	0	1	-4	0	-9
	2	86	75	54	31	26	30	64	-37	-24	-13	-2	0	5	-19
	3	81	67	50	30	29	29	61	-43	-33	-19	-7	-8	-8	-29
Facets & orientation	1	60	55	38	23	24	24	40	-17	-13	1	3	-4	0	-9
	2	56	54	46	27	26	29	43	-11	-13	-12	-2	-4	1	-9
	3	85	69	50	30	30	31	61	-43	-33	-19	-5	-3	-8	-29
	4	67	67	58	34	31	31	46	25	26	28	13	1	10	8
Reflections	1	59	54	42	23	24	24	39	-15	-13	0	1	-4	0	-9
	2	81	67	56	33	32	35	66	-43	-35	-21	-15	-17	-17	-38
	3	58	54	46	25	24	28	47	-11	-4	0	1	-1	2	-8
Albedo, Emissivity	1	60	55	38	23	24	24	40	-17	-13	1	3	-4	0	-9
	2	93	80	56	33	34	37	76	-22	-13	-5	-1	-4	-1	-9
	3	65	61	50	27	26	29	46	-15	-13	-13	-4	-3	1	-12
ΔQ_S calculation	1	60	55	38	23	24	23	47	-17	-20	1	3	-2	0	-18
	2	65	63	50	27	30	29	48	-11	-13	-12	0	-1	4	-12
	3	63	59	47	27	25	26	40	-17	-13	-13	-4	-7	0	-9
Complexity	1	59	54	42	26	24	25	40	-2	-7	0	1	-3	0	-9
	2	85	69	50	26	25	28	61	-38	-32	-17	-5	-3	1	-19
	3	86	75	56	33	30	31	66	-37	-24	-13	-4	-7	1	-19

Table 4.6: As Table 4.4, but for the net storage heat flux (ΔQ_S)

		Mean Obs	41	32	22	13	4	20	21	*MBE*						ΔQ_S
		RMSE	O/N	D/J	F/M	A/M	J/J	A/S	O/N	O/N	D/J	F/M	A/M	J/J	A/S	O/N
Vegetation		1	96	98	85	48	37	55	67	-20	-24	-19	-13	-11	-23	-15
		2	86	89	78	46	35	50	63	-27	-27	-23	-12	-4	-16	-16
		3	89	89	78	52	42	52	59	-28	-28	-20	-12	-5	-18	-16
Temp. Var. Q_F		1	86	89	78	44	34	50	61	-26	-27	-23	-12	-5	-17	-15
		2	81	84	82	54	50	64	61	-28	-29	-22	-37	-33	-40	-25
		3	97	98	85	49	37	57	69	-30	-27	-20	-13	-10	-23	-16
Facets & orientation		1	73	76	65	38	31	42	52	-16	-14	-14	-11	-11	-20	-14
		2	87	91	82	44	34	50	61	-28	-29	-23	-12	-5	-18	-17
		3	96	97	85	54	49	57	69	-13	-22	-22	-11	-10	-23	-16
		4	109	109	97	54	40	58	76	-31	-27	-18	-8	-4	-10	-11
Reflections		1	81	83	71	39	30	42	55	-19	-24	-24	-11	-4	-16	-16
		2	96	97	85	48	36	51	64	-30	-27	-20	-11	-9	-20	-16
		3	91	96	84	50	39	57	63	-28	-29	-22	-13	-5	-17	-15
Albedo, Emissivity		1	73	76	65	38	31	42	52	-16	-14	-14	-11	-11	-20	-14
		2	84	88	75	44	40	51	59	-24	-26	-24	-12	-4	-14	-16
		3	91	93	84	47	35	51	64	-28	-27	-22	-12	-7	-18	-16
ΔQ_S calculation		1	67	67	59	52	44	53	52	-8	-1	-14	-33	-31	-34	-16
		2	89	93	85	54	40	62	66	-28	-27	-22	-11	-7	-21	-18
		3	90	90	79	42	34	50	62	-27	-27	-23	-12	-4	-16	-15
Complexity		1	82	86	73	46	33	49	57	-25	-26	-22	-11	-5	-18	-16
		2	90	89	79	47	35	50	63	-28	-28	-23	-12	-5	-17	-17
		3	99	100	87	49	37	57	68	-28	-27	-20	-8	-10	-21	-13

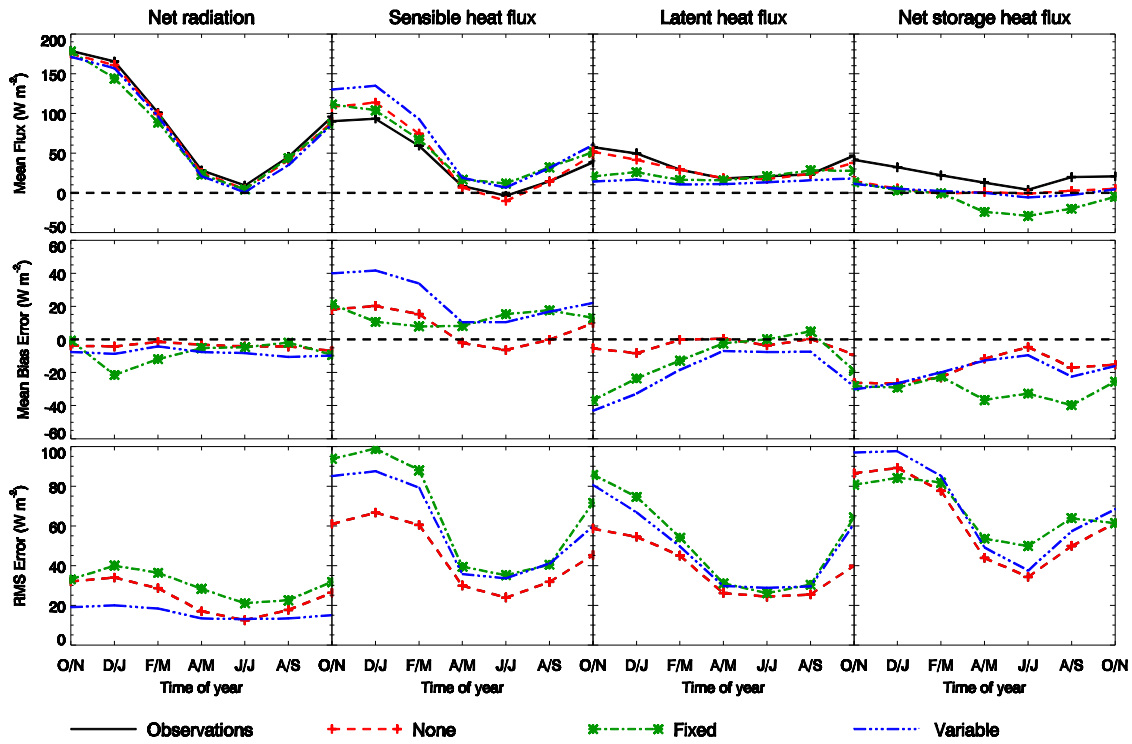

Figure 4.9: Median of the mean modelled flux, mean bias error, and root-mean-square error for the surface fluxes determined for two month periods, for the models classified by the temporal variation in the anthropogenic heat flux

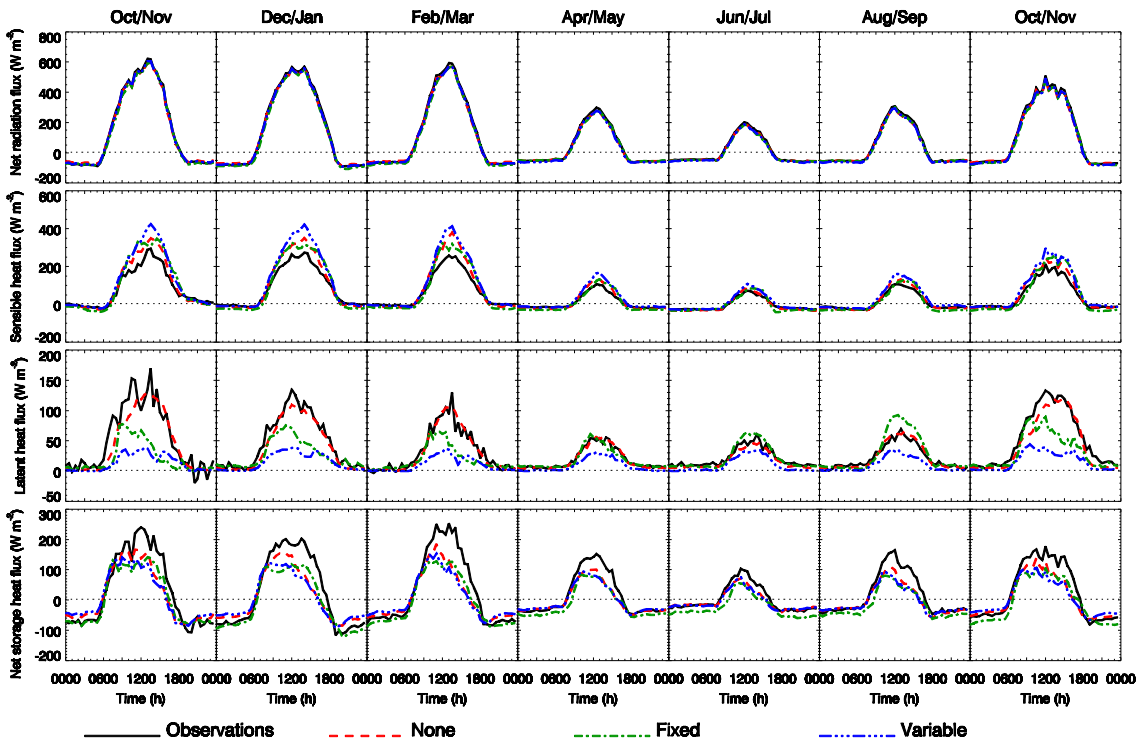

Figure 4.10: Median of the average diurnal cycle for each 60-day period throughout the seasonal cycle, from the models classified by the temporal variation in the anthropogenic heat flux. Note the scales are different for each flux.

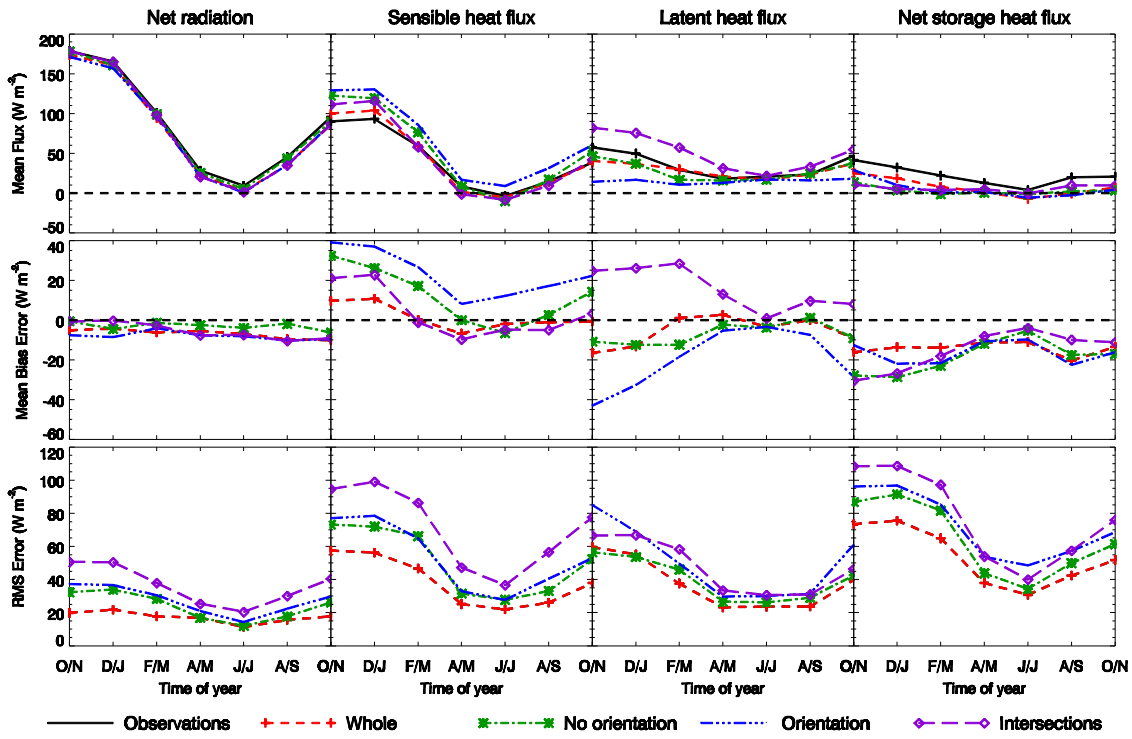

Figure 4.11: As Fig. 4.9, but for models classified by their facets and orientation.

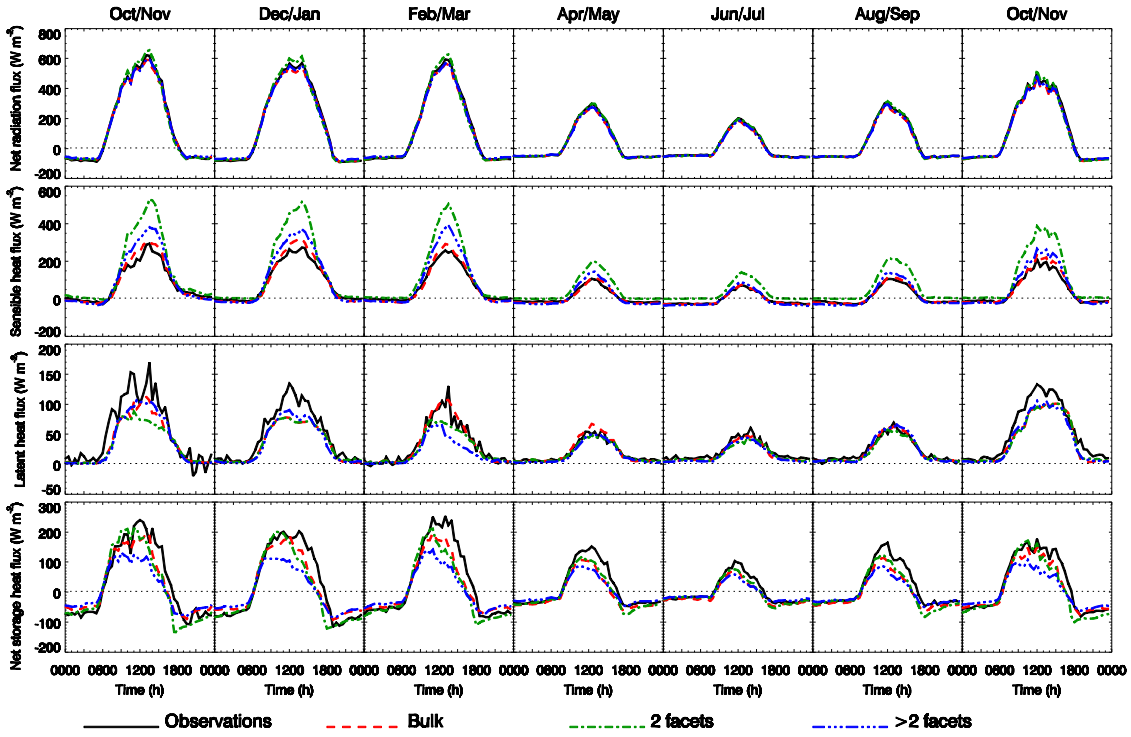

Figure 4.12: As Fig.10, but for models classified by the number of facets with a distinct surface albedo.

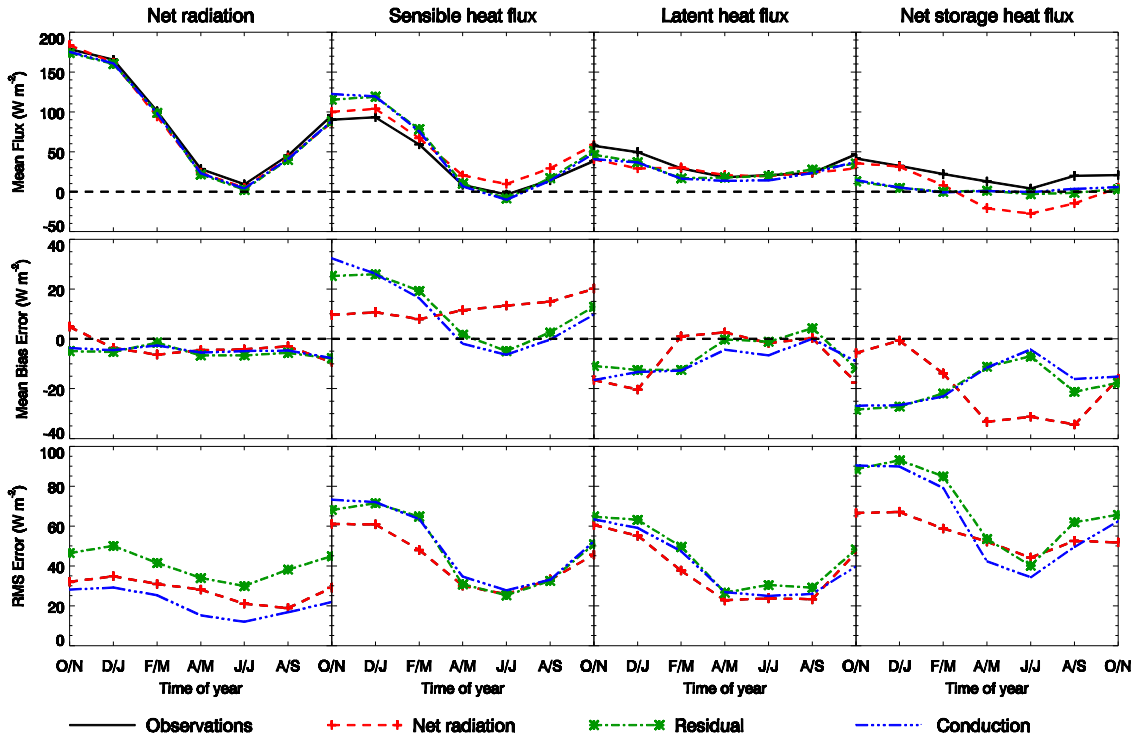

Figure 4.13: As Fig. 4.9, but for models classified by the method for calculating the net storage heat flux.

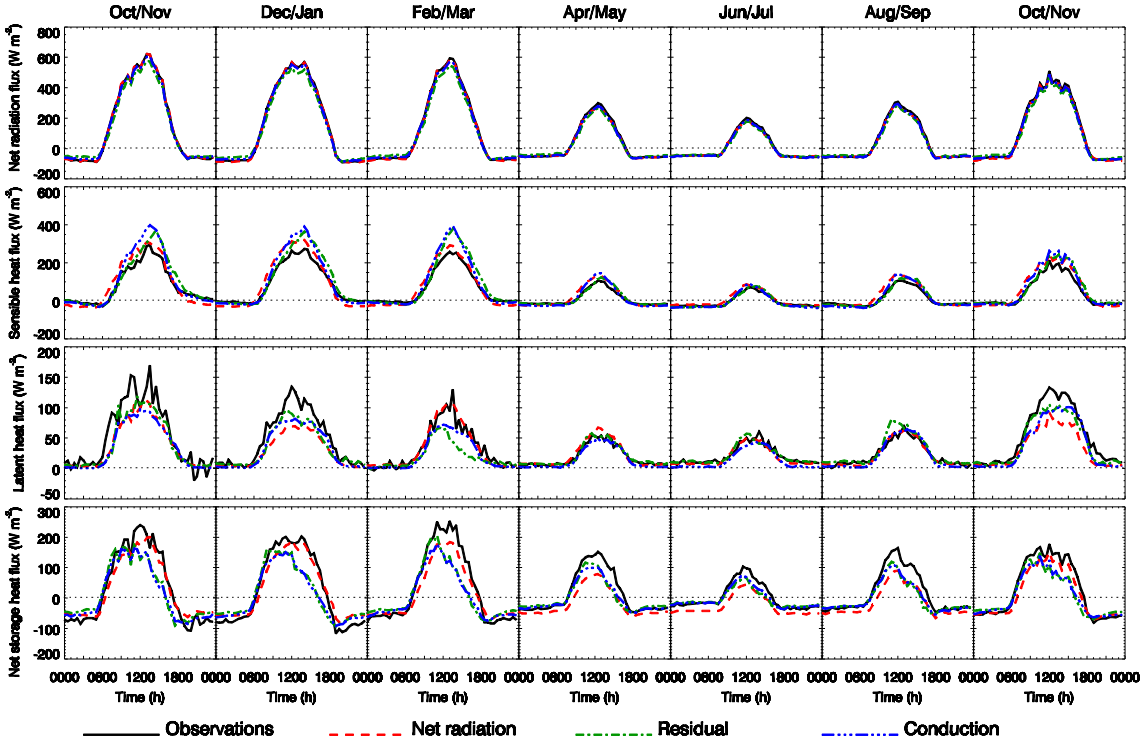

Figure 4.14: As Fig. 4.10, but for models classified by the method for calculating the net storage heat flux.

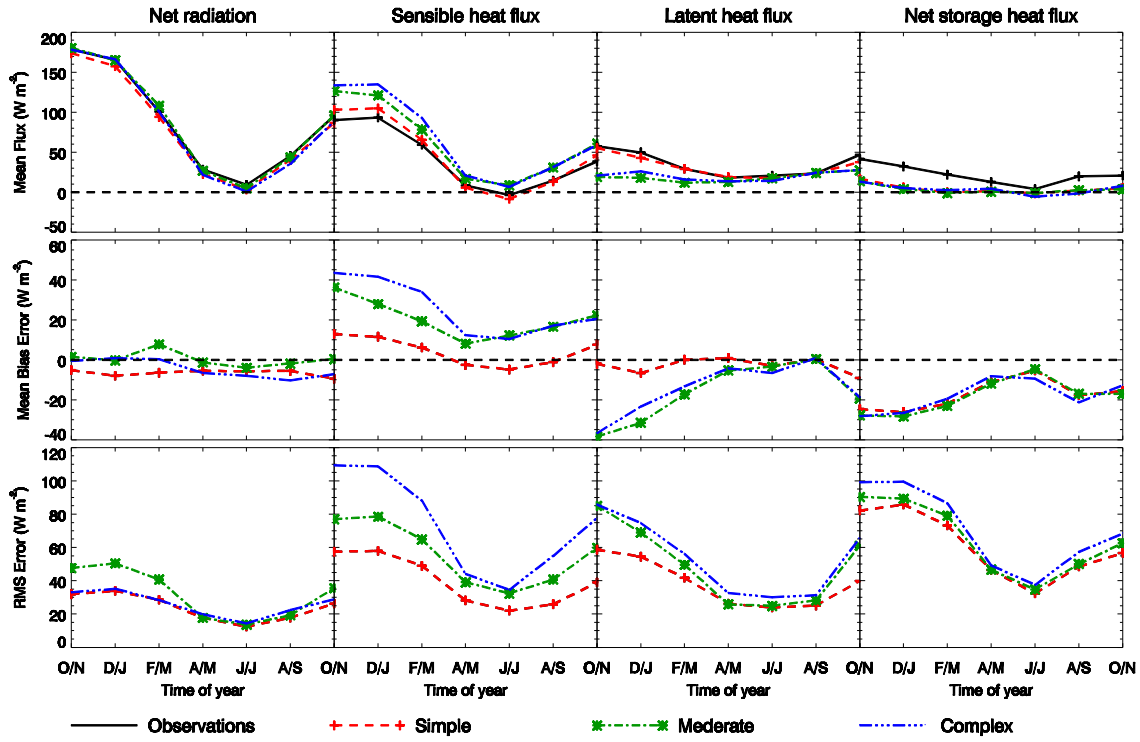

Figure 4.15: As Fig. 4.9, but for models classified by their complexity.

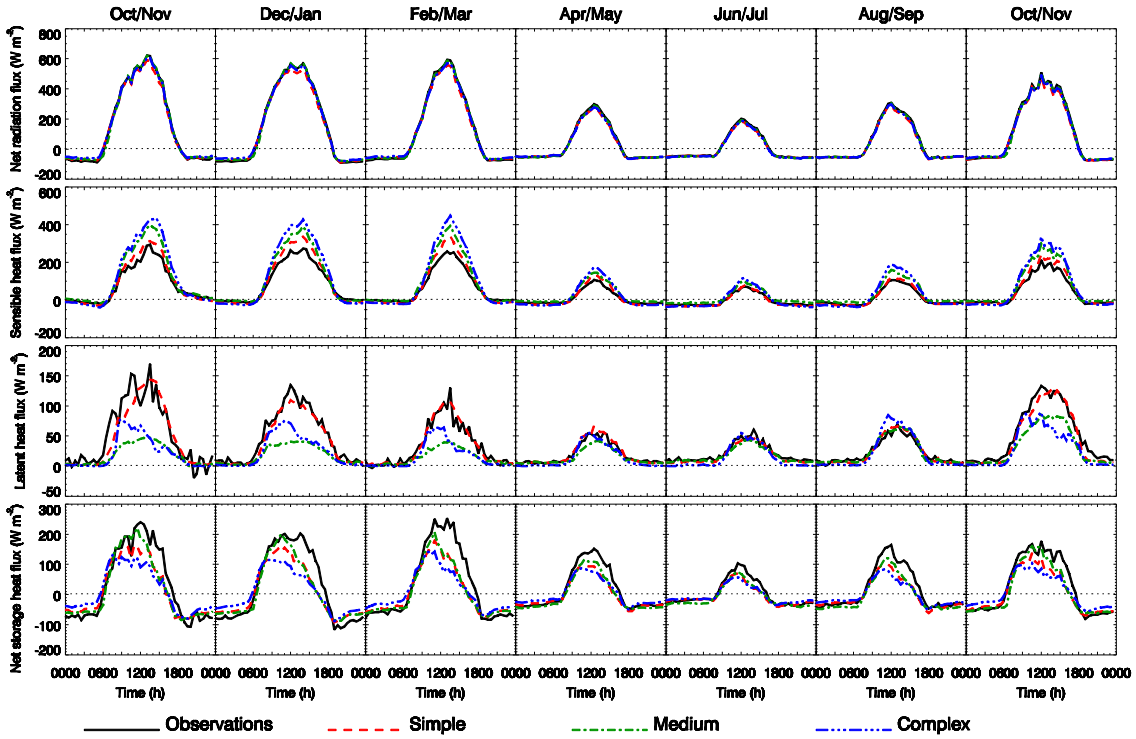

Figure 4.16: As Fig. 4.10, but with models classified by their complexity.

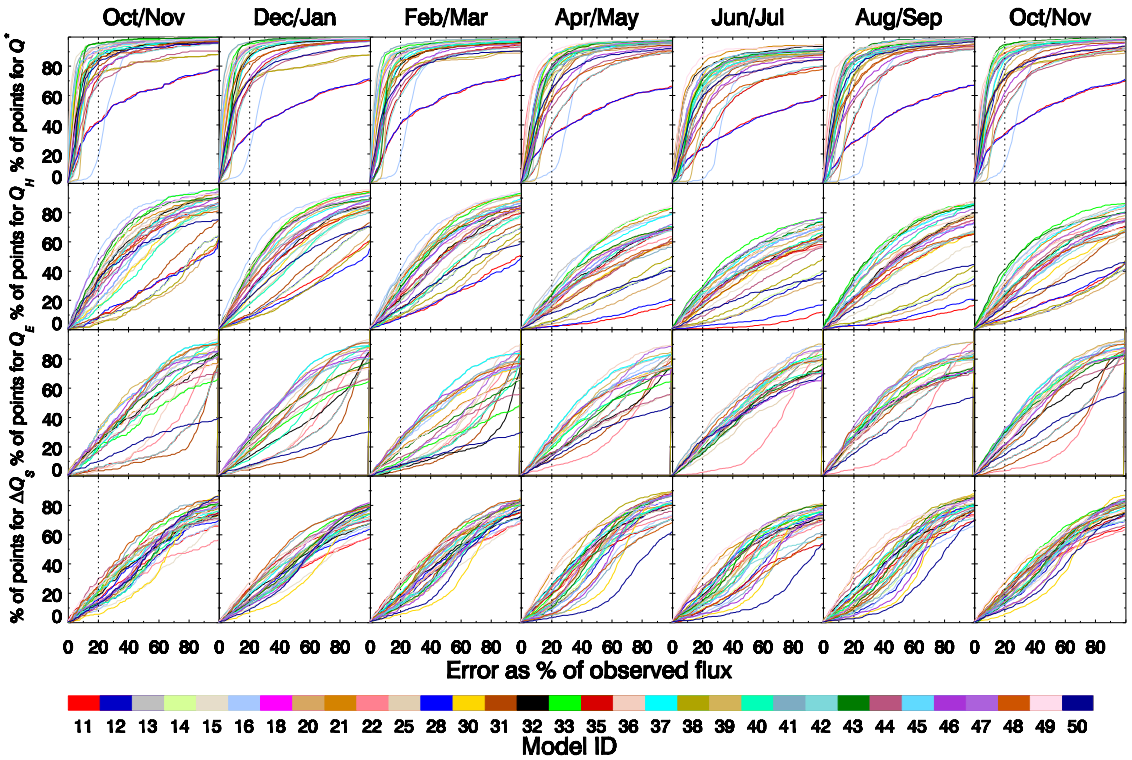

Figure 4.17: Percentage of modelled data within a percentage of the observed flux for each individual model for each flux and 60-day period. An observational error estimate of 20% (the value for the possible under estimation in the sensible and latent heat fluxes as suggested by Coutts et al. 2007b) is indicated by the vertical dotted line.

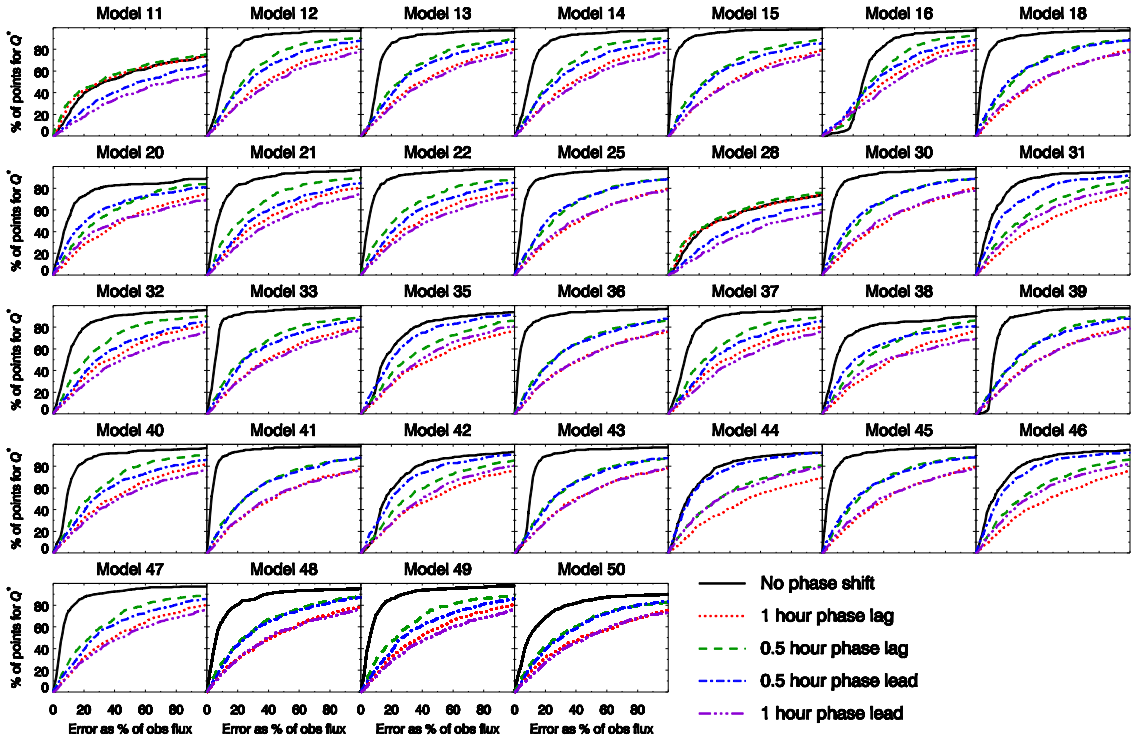

Figure 4.18: Percentage of modelled data within a percentage of the observed flux for each individual model during the February-March period, for the net radiation flux. The modelled results have been shifted relative to those returned by the modelling group: 30 min and 60 min earlier and later.

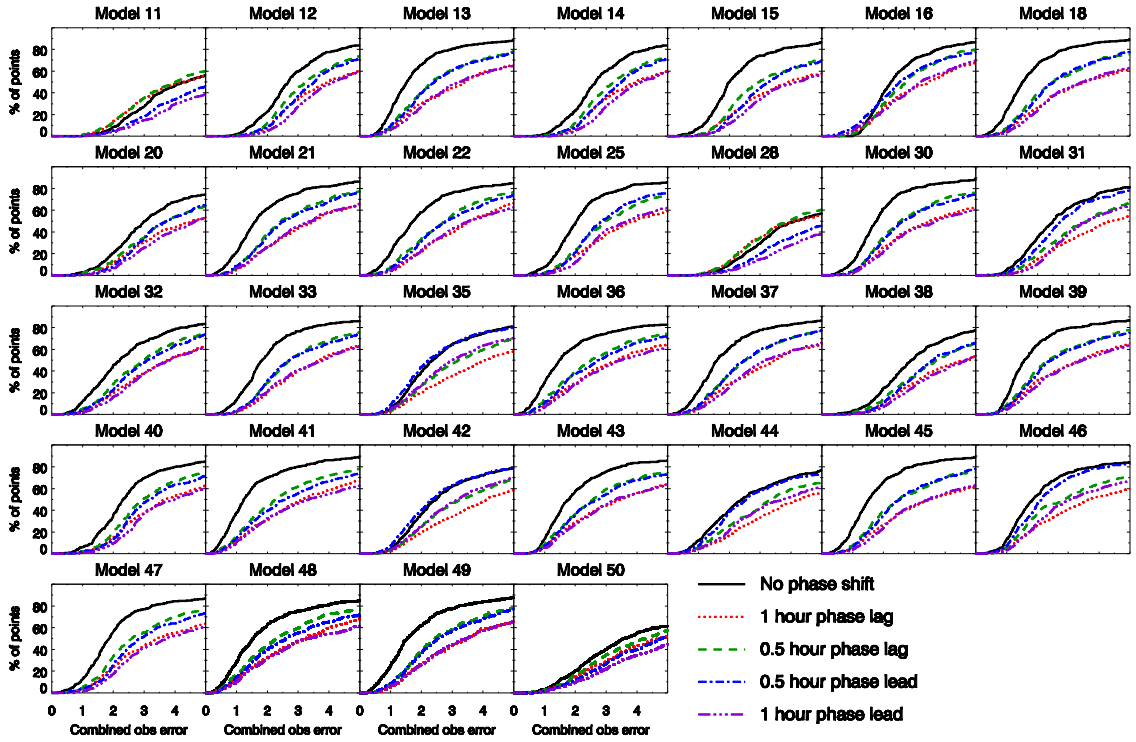

Figure 4.19: As for Fig. 4.18, but for the model data combined using the observational error estimates given by Hollinger and Richardson (2005) as a fraction of the observed net radiation, for the net all-wave radiation (5%), the sensible heat flux (10%) and the latent heat flux (8%). The net storage heat flux is omitted as it is determined from the residual of the energy balance within the observational dataset.

Chapter 5: Importance of initial state and atmospheric conditions for urban land surface models' performance.[11]

Abstract

Urban land surface models (LSM) are commonly evaluated for short periods (a few weeks to months) because of limited observational data. This makes it difficult to distinguish the impact of initial conditions on model performance or to consider the response of a model to a range of possible atmospheric conditions. Drawing on results from the first urban LSM comparison, these two issues are considered. Assessment shows that the initial soil moisture has a substantial impact on the performance. Models initialised with soils that are too dry are not able to adjust their surface sensible and latent heat fluxes to realistic values until there is sufficient rainfall. Models initialised with too wet soils are not able to restrict their evaporation appropriately for periods in excess of a year. This has implications for short term evaluation studies and implies the need for soil moisture measurements to improve data assimilation and model initialisation. In contrast, initial conditions influencing the thermal storage have a much shorter adjustment timescale compared to soil moisture. Most models partition too much of the radiative energy at the surface into the sensible heat flux at the probable expense of the net storage heat flux.

5.1 Introduction

The resolution of weather forecasting models now enables land cover that was previously treated as sub-grid-scale heterogeneity to be explicitly resolved. This has resulted in a number of land surface schemes being developed to explicitly represent urban areas (e.g., Table 5.1). Combined with the increasing interest in the impact of climate change on the urban environment, this has led to the introduction of urban surfaces within the tiled land surface heterogeneity approach for climate models (e.g., Oleson et al. 2008a, McCarthy et al. 2010).

These urban land surface models (LSM) vary considerably in their complexity, from simple bulk representations, to models that divide a wall surface into several layers, each of which has its own surface energy balance (e.g., Table 5.1). To evaluate how these various urban LSM simulate the net all wave radiative flux and turbulent fluxes of sensible and latent heat, the

[11] *This has been published as:* Best MJ, Grimmond CSB (2014) Importance of initial state and atmospheric conditions for urban land surface models performance. Urban Climate, 10, 387-406. doi:10.1016/j.uclim.2013.10.006

models have been compared to data from observational campaigns (e.g., Ross and Oke, 1988, Grimmond and Oke, 1991, 2002, Masson et al., 2002, Best et al., 2006, Dupont and Mestayer, 2006, Hamdi and Schayes, 2007, Krayenhoff and Voogt, 2007, Oleson et al., 2008a, Kawai et al., 2009, Porson et al., 2009, Loridan et al., 2010, 2011, Järvi et al., 2011). Given the difficulties of taking observations within the urban environment, these campaigns have often been of relatively short duration of a few weeks to months (e.g., Grimmond et al., 1996, Oke et al., 1999, Voogt and Grimmond, 2000, Grimmond et al., 2004a, Mestayer et al., 2005, Offerle et al., 2005b, Newton et al., 2007). This has meant that it is not possible to determine the impact of the initial conditions, or the strategy chosen to spin-up the model from these initial conditions, on the performance of models.

Although the initial conditions can impact on a model integration, if the right spin-up strategy is adopted the influence of the initial conditions can be removed. To correctly spin-up the initial conditions, there needs to be a period of sufficient length that the memory of the initial conditions has been lost. For certain parts of the land system, such as deep soil moisture, the length of time required for spin-up can be many years, or even decades (e.g., Chen et al., 2007). For initial conditions to be correct, a dataset that spans multiple years prior to the period of interest are required. As this is usually unavailable, and would severely restrict possible study periods, spin-up strategies are often designed around repeating one given year of data a number of times. However, this strategy brings the model into equilibrium with that particular year rather than providing the correct initial conditions for the study period. As few urban observational datasets are multi-year, it is not possible to adopt this strategy for model spin-up, hence most urban studies have minimal spin-up and rely on their initial conditions being suitable. Thus the initial conditions can impact the evolution of the model, if incorrect values are used.

Many urban LSM were systematically assessed within the first urban LSM comparison (PILPS-urban) (Grimmond et al., 2010, 2011), using an observational period of 474 days. Initial results found a representation of vegetation is critical for accurate simulation of the surface turbulent fluxes of sensible and latent heat. This conclusion is supported when the LSM results are analysed at a higher temporal resolution which resolves the seasonal cycle (Best and Grimmond, 2013). However, it is well known that for vegetated surfaces, the partitioning between sensible and latent heat fluxes is largely determined by the availability of moisture in the soil and that changes to this soil moisture can occur on seasonal to annual timescales, especially for the moisture at deeper soil levels (e.g., Dirmeyer et al., 2006, Notaro et al., 2010). Therefore soil moisture initialisation can impact on the evolution of the sensible and latent heat fluxes for vegetation on these timescales. Hence it is reasonable to assume that soil moisture

will impact the surface energy balance of an urban environment and that soil moisture profile initialisation could be important. Whilst urban models might have other water stores in their system (e.g., precipitation lying on a road or roof surface), the timescale associated with these stores does not compare in length to that of the deep soil moisture.

Urban environments typically have a higher thermal inertia compared to their surrounding rural areas, with net storage heat fluxes that are substantially larger than the comparative soil heat fluxes, causing a time delay in changes to the surface temperature of the urban facets (walls, roof, road) which will impact the energy balance fluxes (Grimmond and Oke, 1999a). Any errors in the initial temperature profile within the surface facets could persist, due to the large thermal inertia, impacting LSM performance. However the models use different representations of the surface, varying in complexity from a bulk representation of the urban area with a single surface temperature, to surface temperatures for a number of facets, to a number of surface temperatures for a single facet (such as several vertical levels for a wall). Thus it is difficult to identify a variable (e.g. surface temperature) that is observable at an appropriate scale, that can be used to evaluate all of the models.

Assessing the impact and timescales that information is retained from initial conditions can be difficult, as often the data required to initialise can not be, or are not, observed. Unfortunately sensitivity studies can not distinguish between initialisation errors and model formulation errors. For instance, Best et al. (2006) demonstrated that the observed latent heat flux for an industrial site in Vancouver, Canada, could only be obtained if the model was initialised with unrealistically high values of soil moisture. Independent of horizontal advection possibly impacting these observation (Masson et al. 2002), it is possible to get realistic fluxes for the wrong reasons, such as unrealistic initial conditions.

So to distinguish between model formulation errors and initial condition, the simulation period needs to be sufficiently long so that the initial conditions are no longer retained while ensuring the seasonal variations in model performance are considered. In addition, observational datasets of short duration are unlikely to capture all of the synoptic weather conditions experienced at the site over each of the seasons, making it difficult to draw conclusions about the suitability of a model to represent an urban region in general. This suggests simulation for at least twelve months is required.

Whilst there are still few urban datasets of this length, the one used for the first international urban model comparison (Coutts et al., 2007a, 2007b) is of sufficient length to allow for both the exploration of the impact of initial conditions on the performance of different urban models,

and the impact of various atmospheric conditions on the performance of the urban models. Hence the objectives of this paper are two fold: first, to analyse urban LSM performance to assess the impact of user choices of initial conditions and the timescales these impacts persist, and second to assess how urban LSM performance changes with different weather conditions.

5.2 Methods

The urban model comparison was designed to provide a systematic evaluation of a number of urban LSMs. The overall goals of the comparison included the identification of: the dominant physical processes within the urban environment, the complexity required for an urban LSM to be suitable to represent the radiative, sensible and latent heat fluxes, and the subsequent parameter requirements for such a model. The model evaluations were performed for a single site using meteorological forcing data that was observed at the site provided to the modelling groups. The amount of information about the site (such as the vegetation fraction, the height to width ratio of the built environment, the buildings' thermal parameters, etc.) released to the modellers was provided in stages, so that analyses of the changes in results between stages could be used to address these goals (see details in Grimmond et al. 2011). To assess the impact of initial conditions "Phase 2/Stage 4" data, when all information about the observation site was provided, are analysed. Ideally the model performance should be its best at this stage, but Grimmond et al. (2011) found this is not the case for all models.

The forcing and evaluation data used were observed over a 16 month period (13 August 2003 to 28 November 2004) in Preston, a suburb in north Melbourne, Australia (Coutts et al., 2007a, 2007b). The calculated land cover within the turbulent fluxes footprint was 44.5% buildings, 13% roads, 4.5% concrete, 15% grass, 22.5% other vegetation and 0.5% bare soil (Coutts et al. 2007a). The Oke (2004) Urban Climate Zone (UCZ) is 5 and the Loridan and Grimmond (2012a,2012b) Urban Zone for Energy exchange (UZE) is medium density (MD). The mean height to width ratio is 0.42 and the mean wall-to-plan ratio is 0.4.

The observations undertaken at 40 m, were analysed at 30 min intervals from samples at 1 Hz for the radiative fluxes and 10 Hz for the eddy covariance data. Further details are given in Coutts et al. (2007a, 2007b), Grimmond et al. (2011) and Best and Grimmond (2013). The anthropogenic heat flux was estimated following Sailor and Lu (2004) using inventory methods with the contributions from vehicles, building sector and human metabolism accounted for. The net storage heat flux, determined as the residual of the surface energy balance, accumulates the observational errors from the other terms in the energy balance.

Missing forcing data were gap filled to provide a continuous 30 min resolution data set. Evaluation is undertaken for only the time periods when all fluxes are available (8865 30-min periods), hence around 40% of the data are usable in this evaluation. To separate the impact of the initial conditions from the overall performance of the models, whilst also considering the seasonal cycle and enough data points per period, the analysis is performed for 60 day periods (Best and Grimmond, 2013). This yields seven 60 day periods, with the first and last being for the same time of the year. Prior to the first 60 day period there are an additional 53 days of forcing data. However, there are insufficient periods when all fluxes are available to allow meaningful statistics to be analysed from the model output. Thus, these 53 days are used as a spin-up period to try to remove the impact of the initial conditions and only the subsequent periods are analysed. No advice was given to the modelling groups to determine how they should set their initial conditions, or if an additional spin-up should be undertaken. These decisions were therefore made by the modelling groups themselves. This length of spin-up is likely to remove the initial conditions for moisture stores such as water on a road surface, but is not sufficient to spin-up the deeper soil moisture stores. Hence the results from the initial conditions impacting on model performance due to water availability are more likely to be connected to soil moisture and vegetation processes than the initial conditions for other water stores.

To assess the impact of the soil moisture initial conditions, the Bowen ratio is compared between the first and last of the 60 day periods. The Bowen ratio is explicitly linked to the soil moisture through the evapotranspiration from the vegetation fraction within the urban footprint. As some models were run by more than one modelling group (Table 5.1) with independent assumptions about the initial conditions, it is possible to separate the impact of the initial conditions from the evaluation of the model physics. The Bowen ratio, for the midday period (10:00 - 14:00 local standard time), is determined from the mean sensible and latent heat fluxes for the period. Similarly for the impact of other initial conditions, such as the temperatures of the various facets that make up the urban surface, the ratio of any two fluxes is determined from the mean value of each flux over the same midday period.

To investigate the impact of atmospheric conditions, the observed and modelled net all-wave radiation (net shortwave plus net longwave), sensible and latent heat fluxes and net storage heat flux have been scaled by the incoming all-wave radiation flux so that a direct comparison can be made between these surface fluxes (Loridan and Grimmond 2012a, 2012b). The median of these results, along with the standard error, has been calculated for a number of bins based upon the value of various atmospheric variables that are used to force the urban models. The size of the

bins for each atmospheric variable have been chosen to allow the distribution of results across the range of the variable, whilst ensuring that there are enough data to create robust statistics for each bin. This results in a different number of bins between variable analysed ranging from eight for temperature to 16 for wind speed.

5.3 Results

5.3.1 Impact of model physics

A number of things influence the performance of each of the urban models; this includes: how the physics within the model are represented, the values of parameters used, the initial conditions of the prognostic variables in the model and any spin-up strategy that is used to reduce the influence of these initial conditions. Within the urban model comparison no spin-up protocol was specified. Therefore the initial conditions of each model could still influence model performance, especially as many urban model evaluation studies have only been able to use repeated use of the initial period rather than using a long period of preceding data due to the short duration of the observational datasets, alternatively spin-up is neglected. Although the choice of model parameters and the representation of the physical processes within the models can both influence performance they are treated here as "model physics" for the comparison to observations; and initial conditions are a second aspect.

The total observed precipitation during the simulation period (all hours independent of availability of observed fluxes for evaluation) was 883 mm, which occurred during 1089 of the 30 min periods. These data were provided within the forcing data set. To assess the net variation in model performance the total cumulative evaporation (over all hours of the simulation) from each of the models are compared (Fig. 5.1a). It is evident that there is a large spread in the results with model 22 having a cumulative evaporative flux of 164 mm and model 44 of 821 mm (whole period). The evaporative flux for the restricted period when all the fluxes were observed was 214 mm (observed) whilst the two models were 52 and 358 mm respectively. These two models both have normalized Bowen ratios of $1.0 \leq \beta_N \leq 1.5$. Although the two models had similar evaporation rates for their first and last 60 day periods, their actual evaporation time series is quite different (Fig. 5.2). Model 22 underestimates the evaporation throughout the year, whereas model 44 has a much better fit to the observed latent heat flux, but in general overestimates during the winter period.

One model (50) has a net evaporation that is greater than the total precipitation over the period

of the simulation (or integration—if not the whole period) and generally over estimates evaporation throughout the year. To sustain these large evaporation rates over an extended period of time would require irrigation. Unfortunately it is unknown which of the models simulated this. This data was not provided as part of the forcing data set and the area was under irrigation restrictions during the study period. So the net loss of water to the system simulated may be unrealistic over the period, and is unlike the results from any of the other models within the comparison (Fig. 5.1a). However it is likely that whilst the model physics has the largest impact on the differences in results, the initial conditions for model 50 are impacting its performance.

As each model does not change its representation of the physical processes during the simulation period, it is possible to identify the impact of the initial conditions on the performance for each of the models by comparing the behaviour of the model during the first and last 60 day period (i.e., the same period in consecutive years) relative to the observed behaviour at the same time. For example, an increase in the mean evaporative fraction (latent heat flux divided by the sum of the turbulent sensible and latent heat fluxes) in the observations over the same period of the year would suggest that there is more water available for evaporation in the second summer compared to the first. As the models are forced with the observed precipitation they should have the same trends as the observations despite any differences in their representation of the physical processes. So any disagreement between the model behaviour and observations is the impact of the initial conditions used for the simulation.

5.3.2 Initial Conditions: Bowen ratios

As the treatment of vegetation within the models plays a key role in their performance (Grimmond et al., 2010, 2011, Best and Grimmond, 2013), and evapotranspiration is influenced by soil moisture availability, the seasonal and annual evolution of soil moisture are important. Therefore one metric of the impact of the initial conditions is the change in the Bowen ratio between the start and end of the simulations.

The observations show that the seasonal cycle of the midday Bowen ratio (β) has an increase over the first 6 months of the summer period followed by a decrease in the winter time (Fig. 5.3a). During the summer months the larger values of net all wave radiation drive larger latent heat fluxes which reduce the soil moisture. This reduction in soil moisture is not sufficiently replenished by the precipitation that is reducing during this period, and hence evapotranspiration becomes limited by the availability of soil moisture. The net result is an increase in the average

Bowen ratio (Fig. 5.3a). The observed Bowen ratio decreases after the larger autumnal rainfall in April and May. Overall, the first late spring (October/November) had a higher β than the following year which was wetter. The observed normalized midday Bowen ratio (first 60 day period normalised by the last 60 day period, β_N) was 1.24; i.e., β was larger (soil drier) at the start of the observation campaign than at the same time the following year (Fig. 5.1b).

The simulated Bowen ratio is influenced by a number of model processes which can include the soil water, the leaf area index of the vegetation, the depth of soil over which vegetation can extract water, the texture of the soils, etc. Thus initial soil moisture is not the only control on the relative performance of the various models but the comparison is informed by the assessment of the same period in two consecutive years for each model. The changes in the Bowen ratio between these two periods, removes differences in model physics within each model. Hence comparing the normalised Bowen ratio of each model against that observed, allows us to assess how the change in moisture availability for the model differs to the observations.

Ideally the models would have the same proportional decrease in their Bowen ratios over the integration period, and hence the same normalised Bowen ratio as that observed. Six models do not simulate any latent heat flux (see Grimmond et al. 2010, 2011 for discussion) therefore they have an ill-defined β. Thus their β_N can not be compared. Of the remaining models, only 12 have a value of $1.0 \leq \beta_N \leq 1.5$ (i.e., close to the observed $\beta_N \pm 0.25$), whilst seven models have a β_N that is less than 1 (Fig. 5.1b). Eight models have $\beta_N > 1.5$, with three of these greater than 2.0 and a further two with $\beta_N > 2.5$. It is likely that this behaviour is caused by the initial conditions for soil moisture chosen by the user. The high β_N are probably related to low values of available water for evaporation at the start of the simulation restricting the latent heat flux and hence increasing the β, whereas the low values of β_N could be due to the initial available water being too high and hence not restricting the evaporation at all.

To demonstrate that the influence of initial conditions can be separated from the representation of model physics, the results for three models with similar cumulative evaporation, but different β_N behaviour (*viz*, models 41, 25 and 16) are shown in Figure 5.2. Although model 41 Bowen ratio ($1.0 \leq \beta_N \leq 1.5$) is similar between the first and last 60 day periods, the model clearly underestimates the evaporation during both summer periods (Fig. 5.2) suggesting the initial conditions are having little if any impact on the evolution of the model.

For model 25 ($\beta_N > 1.5$) the evaporation for the first summer period is underestimated whereas in the second summer period is in better agreement with the observations (Fig. 5.2). This suggests there is an available water limitation during the first summer but subsequently, after

sufficient precipitation the model does not have this as an ongoing constraint so performs more realistically.

With model 16 (β_N < 1.0) the evaporation is more restricted during the second summer period compared than during the first, compared to the observations (Fig. 5.2). Given that the observed Bowen ratio decreases between the first and second summer periods (i.e., there is more water available for evaporation in the second summer that there is during the first), this suggests that the physics of model 16 would generally underestimate the observed evaporation due to water availability. This indicates that the initial conditions have been set so that there is little limitation to the evaporation but through time the available water is depleted, despite receiving precipitation in the winter, so it becomes a constraint.

To consider the persistence and impact of the initial conditions three representative cases (Fig. 5.3a) are chosen: a) initial β more than double the final period value (Model 40); (b) starting β smaller than at the end (Model 50); and (c) similar β_N to the observations (Models 39, 46 with values of 1.22 and 1.23 respectively).

In the first case (40), the high β persists for the first 6 months when there is less seasonal rainfall, but in April/May when the rainfall increases again the β decreases to similar values to that observed (Fig. 5.3a). For the subsequent 60 day periods, the β are more similar to the observations, which suggests that model 40 was initialised with too dry soil moisture that restricted evapotranspiration unlike the observations. It is only once more rainfall occurs during the winter months, when the net radiation is limited, that the modelled soil moisture is able to recover to values which do not limit this evapotranspiration.

Although models 39 and 46 have β_N that are similar to the observations, the actual values of the β are not the same (Fig. 5.3a). Compared to the observations, model 39 has a similar shape to the seasonal cycle of β but there is an off-set throughout the simulation, with the model having higher values than the observations. The β values for model 46 compare well with the observations during the winter period, but have larger vales than observed in the autumn and both of the summer periods.

In the third case, model 50 initially does not follow the observed increase (December/January) but does start to increase in February/March (Fig. 5.3a). However, the β continues to increase in the autumn and winter months, before reducing in the spring. This may be caused by initial soil moisture that is too high. The initial reduction in soil moisture during the summer months is not sufficient to limit latent heat fluxes. However, the soil moisture continues to decrease during the

autumn and winter which then restricts the evaporation and hence β increases, even as the rainfall start to increase. It is only once the accumulated rainfall reaches its larger values that the soil moisture is increased to a level that reverses this trend.

The differences in the evolution of the β between model 50 and others is not driven solely by the initial condition of soil moisture, but also by the amount of soil water that is removed via evapotranspiration. The average mean bias error (*MBE*) in the latent heat flux from model 50 is around double that from any of the other models with a positive bias (Table 5.2), and so it removes more soil moisture regardless of the initial conditions. Hence although the initial conditions of soil moisture are important, so is the relation between soil moisture and the limitation in evapotranspiration.

5.3.3 Initial Conditions: Model user assumptions

For two of the models in the comparison, results from identical configurations (including physical setup) were submitted by more than one group. This ensured that the only differences between the models were the initial conditions, ensuring a clearer analysis of the role of initial conditions on the evolution of the model fluxes, whilst removing differences due to the physical parameterisations. Fig. 5.3b shows the evolution of the β for these two models (now called "1", "2"). For both models, the simulations consist of results with an initial β substantially larger than the observed value ("dry") and a second set with β results in better agreement with the observations ("wet"). Although the timing of the reduction in the large β varies between the models, both simulations with initially large β have reduced values through the winter/spring months, when the accumulated rainfall is higher. By the end of the simulation, all runs of the models have similar β which are close to the observed value. However, one of the model simulations (1_{wet}) has average midday β that is consistently lower than the observed value.

At the start 2_{wet} is closest to the observed β but there is substantial increase in the modelled β during February/March resulting in values that are close to the initialised dry run (2_{dry}) (Fig. 5.3b). This suggests that the soil moisture dries out to similar values despite the different initial conditions. However, this drying of soil moisture with subsequent increase in Bowen ratio is not seen in the observations. The initially wetter run (2_{wet}) stays drier during the period of larger accumulated rainfall compared to the initially drier model run (2_{dry}), suggesting that the representation of physics within the model is also sensitive to the evolution from its initial state and not just the initial conditions.

The impact of the initial conditions on 1_{dry} and 1_{wet} simulations can clearly be seen in Figure 5.3b. Whilst the two simulations show similar values of evaporation during the initial spin-up period (first 53 days), the evaporation from the 1_{dry} simulation is clearly restricted during the first summer period. This is caused by the onset of a restriction in water for evaporation that is due to the drier initial conditions set within this simulation. However, during the second summer period the evaporation rates are more similar between the two simulations, although 1_{wet} still has larger values than $1_{dry.}$ Hence although the impact of the initial conditions has been reduced by the second year, they have not be removed completely.

5.3.4 Initial Conditions: Sensible heat ratios

As building materials have high thermal inertia, the initial conditions for the temperature of the urban facets (walls, roof, road) could impact the simulations. In addition, although the soil thermal capacities are smaller than that for many building materials, the initial conditions for the soil temperatures in the models may also contribute. The net storage heat flux from the models does not distinguish between urban facets and the underlying soil, therefore these are considered together. If an impact from the initial surface and soil temperatures does exist, then we would expect a trend in the ratio of the net storage heat flux to net all wave radiation ($\Delta Q_S / Q^*$) between the start and end of the simulation period, and a reverse trend in turbulent sensible heat flux to net all wave ratio (Q_H / Q^*). Unlike β, there are no such trends in modelled ratios of the mean midday fluxes (e.g., Fig. 5.4), with the exception of Q_H / Q^* from model 50. There are some small trends in Q_H / Q^* in some cases (e.g., model 40, Fig. 5.4c) but not in $\Delta Q_S / Q^*$, even for model 50. Hence it is not obvious that the initial conditions chosen by the user for the soil temperatures or the surface temperature of each facet in their model have a significant impact on the simulation.

Although there are no obvious differences in the values of Q_H / Q^* and $\Delta Q_S / Q^*$ between the start and end periods, only five models (21, 22, 33, 44 and 49) have values of Q_H / Q^* which are consistently smaller than observed during the seasonal cycle (e.g. Fig. 5.4b). With the exception of model 22 which has a specified constant β (=1), these models all have an initial β that is lower than observed, and this persists throughout the seasonal cycle (not shown). The only other model with an initial β that is smaller than observed is model 50, which shows a trend in Q_H / Q^* over the year. Hence it is likely that these models were initialised with too high soil moisture. Also, apart from model 50, these models have values of Q_H / Q^* that are closer to the observed values by the end of the simulation than at the start. Hence this demonstrates that the initial conditions of soil moisture also impact on the sensible heat flux, through the surface energy

balance, as might be expected. The implications of the initial conditions for the soil moisture are not so clear for the values of $\Delta Q_S/Q^*$ (e.g., Fig. 5.4b), although three of these models (21, 44 and 49) have values which are closer to those observed over the whole period than the other models.

There are a further four models (31, 32, 36 and 40) that have values of Q_H/Q^* which are lower than observed in the winter and spring period (e.g., Fig. 5.4c). Of these four models, three have an initial Bowen ratio that is substantially larger than observed, but are in much better agreement by the end of the simulation (not shown). However, these are not the only models that have an initial Bowen ratio that is substantially larger than that observed and hence the initial conditions on their own are not sufficient to explain the seasonal behaviour of Q_H/Q^* in these models. For the remaining majority of models the ratio of Q_H/Q^* is overestimated at all times of the year (e.g., Fig.4a), whilst the ratio of $\Delta Q_S/Q^*$ is underestimated, compared to the observations. This suggests that in general, the models put too much of the available net radiation at the surface into the sensible heat flux at the expense of the net storage heat flux at all times of the year. This result does not seem to depend upon the initial conditions, unless the soil moisture is set to values which are too high, resulting in a β that is lower than observed throughout the simulation.

This result can also be seen in the midday *MBE* for each of the fluxes (Table 5.2). For the majority of the models there is an overestimate in the sensible heat flux and an underestimate in both the latent heat flux and the net storage heat flux, including models 11 and 28 which have a large positive bias in the net radiation during the daytime. The average *MBE* for these fluxes are 50, -19 and -50 W m^{-2}, respectively; whilst the median model's *MBE* are 42, -11 and -50 W m^{-2}. Exceptions to this include some models that have a positive *MBE* in the latent heat flux, and consequently a negative *MBE* in the sensible heat flux (models 21, 33, 44 and 49). Model 50 has a positive bias in both the sensible and latent heat fluxes due to its excessive evaporation, whilst model 16 has a negative *MBE* for all fluxes, due to its large bias in the net radiation. Only models 22, 36 and 44 have positive *MBE* for the net storage heat flux in the afternoon, with all three models having a negative bias in the sensible heat flux.

5.3.5 Initial Conditions: Surface flux behaviour group by normalised Bowen ratio

To assess the change in performance of the models between the first and last 60 day periods, a Taylor plot (Taylor, 2001) is used for each flux (Fig. 5.5), in order to identify changes to three statistical measures within the same plot. Each statistic shown in the Taylor plot has been

determined using the 30 min data for all models within each β_N category (section 3.1). The group of models with no latent heat flux are not impacted by the initial soil moisture conditions. This means that the change in performance between the first and last 60 day period for this group results from the differences in the climatological conditions and the model's response to these. As such, the results for the no evaporation models can be used as a baseline to assess the impact of the initial conditions for the other groups of models. This is done by comparing the difference between the first and last 60 day periods for each group of models, to that from the group of models without evaporation. Note that models with no evaporation have no vegetation, but a model with no vegetation can still have evaporation if this process is included for other facets of the urban surface.

For net radiation, the best predicted flux, the models that have no latent heat flux have degradation in the root mean squared error (*RMSE*) and correlation at the end of the period compared to the start, although the normalised standard deviation remains about the same (Fig. 5.5a). This is in contrast to all of the other groups of models which show no real difference between the two periods in any of the statistics. This implies that the models generally show an improved performance in the net radiation at the end of the simulation compared to the start, due to the impact of initial conditions of soil moisture, that offsets the degradation that results from the change in meteorological conditions.

For the turbulent sensible heat flux, the models that have no evaporation have a slight improvement in the *RMSE* and normalised standard deviation, but a slightly poorer correlation at the end of the simulation (Fig. 5.5b). The models with a $\beta_N < 1$ have degraded performance at the end of the period for all three of the statistical measures. However, these models have a substantial improvement in the latent heat flux for both the *RMSE* and correlation and an improvement to the normalised standard deviation resulting in good agreement with the observations at the end of the period (Fig. 5.5c). This is the only category of model that has a standard deviation in the latent heat flux for the first 60 day period which is larger than observed.

The models with $\beta_N > 1.5$ have a larger degradation in correlation between the last and first 60 day periods for the sensible heat flux compared to the models with no latent heat flux, but have improvements to the *RMSE* and particularly the normalised standard deviation that are larger than the models with no evaporation (Fig.5b). Those models in the $1.0 \leq \beta_N \leq 1.5$ class show an improvement in all of the statistics at the end of the period compared to the start. However, both β_N classes, plus the overall results for all models, have improved results in all of the statistical metrics for the latent heat flux at the end of the simulation compared to the start (Fig. 5.5c).

Notably the models with $\beta_N > 1.5$ have substantial improvements to the normalised standard deviation for both the sensible and latent heat flux (Figs. 5b, c). These models, initialised with a dry soil state, have a standard deviation that is too large for the sensible heat flux and too small for the latent heat flux during the first 60 day period. This is the expected behaviour if the vegetation was in an incorrectly water stressed state limiting the latent heat flux, resulting in too much energy partitioned into the sensible heat flux. The standard deviations for the last 60 day period do not agree with the observations, but are substantially closer for both fluxes.

All the model classes have an improved net storage heat flux at the end of the simulation compared to the start (Fig. 5.5d). There are some slight differences between the categories compared to the changes in the models with no evaporation, such as a larger improvement to the *RMSE* from the models with $\beta_N > 1.5$, but these differences are smaller than for the sensible and latent heat fluxes. Hence this shows that the initial conditions have more of an impact on the sensible and latent heat fluxes than on the net storage heat flux in general, due to the large influence that water availability has on the evaporative fraction.

For models that are influenced by soil moisture, there is a general improvement at the end of the simulation compared to the start for all fluxes. Whilst the precipitation over the integration period will influence these results, all models received the same precipitation input, so different changes in the errors for the various groups of models show that the initial conditions also contribute. The only exception to this is the sensible heat flux for the models that are initialised with too much soil moisture ($\beta_N < 1.0$, Fig. 5.5b). The largest improvements between the start and end of the simulation period are seen in the standard deviation of the latent heat flux for models initialised with either too much or too little soil moisture ($\beta_N < 1.0$ and $\beta_N > 1.0$ respectively). The models with too much initial soil moisture have more variability in the latent heat flux because it is not sufficiently constrained, whereas those with too little initial soil moisture have less variability due to an over constraint on the evaporation.

5.3.6 Meteorological forcing

As well as the initial conditions, it is important to know the effect of the meteorological forcing conditions on urban land surface model performance. To investigate this, the 30-min model results and observations are analysed across the range of each forcing variables (Fig. 5.6). The fluxes are scaled by the forcing incoming radiation ($Q\downarrow = K\downarrow + L\downarrow$, shortwave and longwave, respectively), with the mean and standard error shown for the bins from each meteorological

forcing. These meteorological forcing (namely shortwave radiation, longwave radiation, rainfall, atmospheric temperature and specific humidity, wind speed and surface pressure) are the data that were given to each group in order to perform their integrations. As such, this data were consistent across all of the modelling groups. In addition, the models have also been analysed for the wind direction, which was not a meteorological forcing that was given to the modelling groups.

Whilst this analysis is not done according to particular weather types certain synoptic situations can be inferred, such as cyclonic conditions having lower shortwave radiation, higher longwave radiation, larger rainfall, higher wind speed and low pressures, or anticyclonic conditions having higher shortwave radiation, lower longwave radiation, less rainfall, higher temperature, lower wind speeds and high pressure.

The normalized Q_H and ΔQ_S observations have a similar magnitude response to atmospheric forcing variables, whereas Q_E is smaller. Figure 5.6 shows an example of this for a model with behaviour that is typical of the majority of models in the comparison. Q^* has the largest response to changes in variations of the forcing meteorological conditions. The almost linear trend between Q^* and the atmospheric temperature is probably linked directly to the radiative forcing of the surface and boundary layer (e.g., model 18, Fig. 5.6).

For both Q^* and Q_H the relation with the humidity forcing (Fig. 5.6) has two maxima and minima. Low values of specific humidity are most likely related to hot, dry anticyclonic periods which have large insolation; whereas high specific humidity are linked to periods associated with precipitation events when the surface water is freely available, such as low pressure systems. Hence the two extremes of the humidity values lead to different behaviour in the net upward radiation and subsequently with sensible heat flux. For the wind speed, the observations have lower values of Q* and Q_H and higher values of Q_E associated with high wind speeds and low pressure (Fig. 5.6), which is also consistent meteorological condition for low pressure systems. These synoptic conditions are likely associated with high humidity and rainfall.

In general the models capture the observed response to each of the atmospheric forcings (e.g., model 18, Fig. 5.6), including the non-linear behaviour with respect to the humidity and wind speed, and the double maximum and minimum for the specific humidity, which agrees well with the observations. However, the models do have a growing bias for increasing shortwave radiation forcing, with Q_H having an increasing positive bias and ΔQ_S with an increasing negative bias. The results for ΔQ_S show that the models have a flat response at larger values of the downward shortwave radiation forcing.

Whilst the majority of models reproduce the observed behaviour for each of the atmospheric forcing variables, the following have differing characteristics (Fig. 5.7). For Q_H, model 11 has higher values than observed for small and large downward shortwave and downward longwave radiation forcing (Fig. 5.7). The values are too high for small and especially large rainfall amounts, and there is a consistent positive bias to atmospheric temperature forcing. There are larger values than observed for lower atmospheric humidity, and an increasing positive bias for increasing wind speed. Finally, there are higher values than observed for low pressure. These results are all consistent with over estimating Q_H during low pressure synoptic storms.

For model 22, with a constant Bowen ratio specified, the results for Q_H do not capture the observed increase with larger downward shortwave radiation forcing (Fig. 5.7), whilst the trend for the downward longwave radiation is the reverse to that observed, apart from largest forcing values. There is a small response to atmospheric temperature, but the inverse to the observed trend, and a flat response to wind speed apart from a reduction at high wind speeds.

The Q_H for model 44 is underestimated compared to observations under large downward long wave radiative forcing (Fig. 5.7). During warmer temperatures the model results have the reverse trend to that observed, and decrease too rapidly with increasing wind speeds.

For model 50, the results for Q_H are too high compared to observations for downward shortwave radiation forcing (Fig. 5.7), whilst the modelled increasing trend in response to the downward long wave radiation between 300 and 400 W m^{-2} is not observed. The modelled values are too large for high rainfall amounts, whereas the increasing trend for wind speed between 5 and 15 m s^{-1} is not observed.

So the majority of the models can simulate the response of the observed surface fluxes to various atmospheric forcing conditions, but there is no obvious consistency between the minority that does not. These few models each have a different response which must be related to the details of their individual parameterisation schemes, rather than the representation of a generic physical process which they share with models in the majority that agree well with the observations.

5.4 Conclusions

For urban LSMs soil moisture initialisation can have a substantial impact on seasonal

performance. If set too dry, urban LSMs are not able to adjust to realistic values until there is sufficient rainfall. Conversely, if initially too wet the model is not able to restrict the availability of water to vegetation. This may persist for a period that is longer than a year, but with smaller errors. This impact of the initial soil moisture state confirms the importance of vegetation and how this interacts with the available water store in the soil of urban areas.

The initialisation of soil moisture has implications for model evaluation studies. As many urban observational campaigns to date have been for short time periods (typically of the order of a few months), without well determined initial soil moisture conditions, evaluation of model performance may be hampered. Unfortunately, urban soil moisture data are still relatively rare. Thus, extreme care should be used to interpret model evaluation results if this is not explicitly addressed. For the modelling community to enhance their model evaluations and their model performance, future observational campaigns of fluxes in urban areas should be complemented with soil moisture observations.

The long term memory from soil moisture is more significant in the evolution of the simulations than for thermal storage, with the timescale for the adjustment of the initial conditions of surface temperatures being much shorter than for soil moisture and not apparent at the end of the simulation. However, the initial soil moisture can also impact on the sensible heat flux and the net storage heat flux throughout the simulation.

In general the models capture the observed trends in the surface fluxes for each of the atmospheric forcing variables, including the observed non-linear behaviour with respect to humidity and wind speed. However, the models are not able to capture the observed increasing trend in ΔQ_S with the larger downward shortwave radiation values. Overall, for increasing downward shortwave radiation, the models put too much energy into Q_H and too little into ΔQ_S. This is consistent with results from analysing the impact of initial conditions (section 3.3). Correcting such behaviour in the models requires future research, as it is not straight forward to increase ΔQ_S without having a similar impact on the turbulent sensible heat flux, due to their common interaction with the surface temperature which forms part of the thermal gradients that drive each of the fluxes.

Some models do not capture the observed relations with atmospheric forcing. However, no general conclusions can be drawn as these models do not share any common representation of their physical processes which implies that the results are not due to the representation itself. Rather it is probably the way in which these processes have been implemented within the parametrisations.

Table 5.1: Models (name, versions (V) and user groups (G)) in Phase 2/Stage 4 of the first international urban model comparison. Model results are presented with anonymous numerical code.

Model name	V/G	References
Building effect parameterization (BEP)	1/1	Martilli et al. (2002)
BEP coupled with building energy model	1/1	Martilli et al. (2002); Salamanca et al. (2009, 2010) ; Salamanca and Martilli (2010)
Community Land Model – urban (CLM-urban)	1/1	Oleson et al. (2008a, 2008b)
Institute of Industrial Science urban canopy model	1/1	Kawamoto and Ooka (2006, 2009a, 2009b)
Joint UK land environment simulator (JULES)	4/2	Essery et al. (2003); Best (2005); Best et al. (2006); Best et al. (2011)
Local-scale urban meteorological parameterization scheme (LUMPS)	2/1	Grimmond and Oke (2002); Offerle et al. (2003); Loridan et al. (2011)
Met Office Reading urban surface exchange scheme (MORUSES)	3/1	Harman et al. (2004a, 2004b); Porson et al. (2010)
Multi-layer urban canopy model	1/1	Kondo and Liu (1998); Kondo et al. (2005)
Nanjing University urban canopy model-single layer	1/1	Masson (2000); Kusaka et al. (2001)
National and Kapodistrian University of Athens model	1/1	Dandou et al. (2005)
Noah land surface model/single-layer urban canopy model	3/3	Kusaka et al. (2001); Chen et al. (2004); Loridan et al. (2010)
Seoul National University urban canopy model	1/1	Ryu et al. (2011)
Simple urban energy balance model for mesoscale simulation	1/1	Kanda et al. (2005a, 2005b); Kawai et al. (2007, 2009)
Single column Reading urban model tile version	4/1	Harman and Belcher (2006)
Slab urban energy balance model	1/1	Fortuniak (2003); Fortuniak et al. (2004, 2005)
Soil model for submesoscales (urbanized)	1/1	Duport and Mestayer (2006); Dupont et al. (2006)
Temperatures of urban facets (TUF) 2D	1/1	Krayenhoff and Voogt (2007)
Temperatures of urban facets (TUF) 3D	1/1	Krayenhoff and Voogt (2007)
Town energy balance (TEB)	1/1	Masson (2000); Masson et al. (2002); Lemonsu et al. (2004); Pigeon et al. (2008)
Town energy balance (TEB) with multi-layer option	1/1	Hamdi and Masson (2008)
Vegetated urban canopy model	1/1	Lee and Park (2008)

Table 5.2: Mean bias error (W m^{-2}) for each flux over the day (24 h) and for midday period (10:00 – 14:00), for each model and for the average and median of all of the models, over the seven 60 day periods. (day 54 to day 473)

Model ID	Q^*		Q_H		Q_E		ΔQ_S	
	24 h	Midday	24 h	Midday	24 h	Midday	24 h	Midday
11	10.5	110.7	70.2	213.2	-34.5	-80.7	-35.8	-33.6
12	-9.6	-37.5	34.9	119.5	-34.5	-80.7	-20.5	-88.1
13	-13.7	-36.6	0.6	20.2	-7.1	-18.8	-17.7	-49.7
14	-9.0	-34.7	35.5	118.6	-34.5	-80.7	-20.5	-84.3
15	-7.7	-9.1	0.5	88.2	16.2	34.5	-13.0	-96.1
16	-23.0	-92.9	-10.5	-0.4	-3.2	-10.8	-19.9	-93.5
18	2.3	-8.3	10.0	60.5	0.5	0.5	-18.8	-81.0
20	-4.7	13.6	55.5	131.2	-34.5	-80.7	-36.2	-48.7
21	-7.0	-18.2	-5.4	-30.1	4.7	12.8	-16.7	-12.7
22	-4.6	2.8	-0.7	-41.6	-26.2	-59.6	11.8	92.3
25	0.3	-13.1	25.8	72.4	-8.8	-10.0	-17.1	-73.5
28	18.7	113.4	59.3	213.8	-34.5	-80.7	-16.6	-31.4
30	3.2	-14.1	12.2	79.6	0.4	0.6	-19.8	-106.0
31	-2.6	-17.5	3.9	39.1	-11.7	-19.9	-17.2	-80.3
32	-0.8	-20.6	11.2	19.5	-2.2	-0.3	-20.3	-51.5
33	-6.2	-21.7	-12.4	-25.7	10.0	35.9	-14.4	-45.7
35	-27.8	-47.9	13.6	1.5	-19.2	-52.4	-32.6	-8.8
36	-8.6	-5.1	19.2	-1.1	-16.5	-33.2	-14.7	27.2
37	-13.8	-35.6	0.4	18.5	-7.1	-18.6	-17.6	-47.2
38	3.1	16.2	44.7	127.4	-34.5	-80.7	-17.6	-42.3
39	-14.7	-45.5	1.8	42.3	-9.2	-20.4	-17.7	-79.2
40	-3.4	-27.0	13.2	53.2	-8.1	-9.3	-18.8	-82.1
41	-10.9	-13.0	8.4	35.6	-9.8	-15.8	-20.0	-44.4
42	-29.8	-48.3	20.7	5.8	-19.2	-52.5	-41.7	-13.3
43	-10.1	-31.2	1.7	21.9	-3.1	7.6	-19.1	-72.5
44	3.7	11.6	-24.4	-44.4	23.4	36.3	-5.8	8.0
45	2.3	-8.3	10.0	60.5	0.5	0.5	-18.8	-81.0
46	-4.4	-20.7	9.0	53.8	-2.9	-0.5	-21.0	-85.9
47	-8.6	-18.9	11.1	41.8	-8.8	-9.6	-21.4	-62.9
48	5.4	-7.6	16.7	38.9	-7.6	-15.1	-14.1	-43.1
49	-6.9	-16.9	-6.6	-32.5	5.6	13.4	-16.4	-9.5
50	-6.6	-8.2	10.6	88.9	38.4	83.9	-38.5	-87.7
Average	-5.8	-12.2	13.8	49.7	-8.7	-18.9	-19.6	-50.3
Median	-6.2	-16.9	10.6	41.8	-7.6	-10.8	-18.8	-49.7

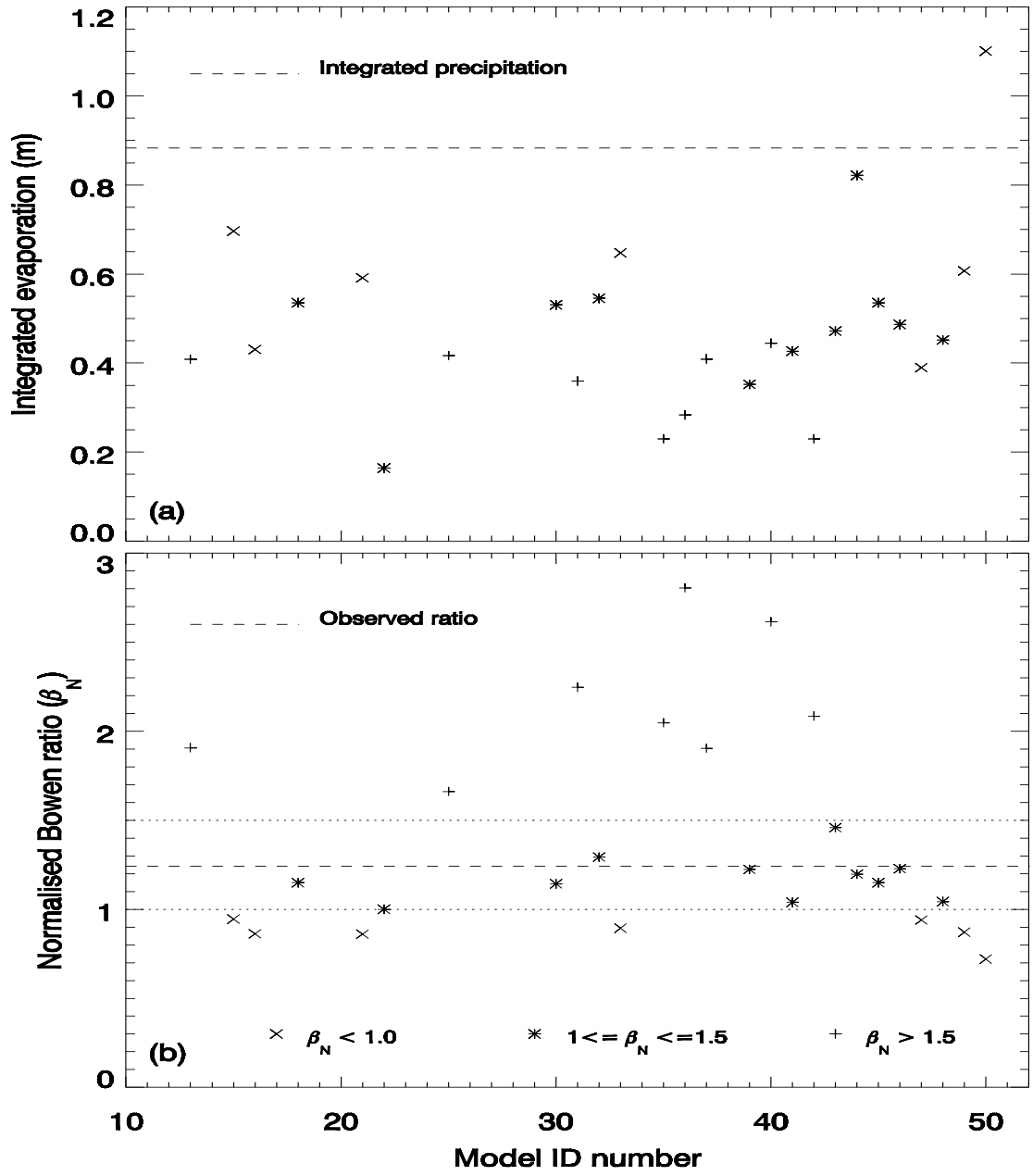

Figure 5.1: Individual model values for (a) total evaporation (m) simulated for the period and total cumulative precipiptation observed (dashed line), and (b) normalised Bowen ratio (first 60 day period normalised by the last 60 day period) relative to the observations (dashed line). To preserve anonymity each model is identified by a random number (between 10 and 50). Symbols in both plots are based on the normalized Bowen ratio (see key) and dotted lines in (b).

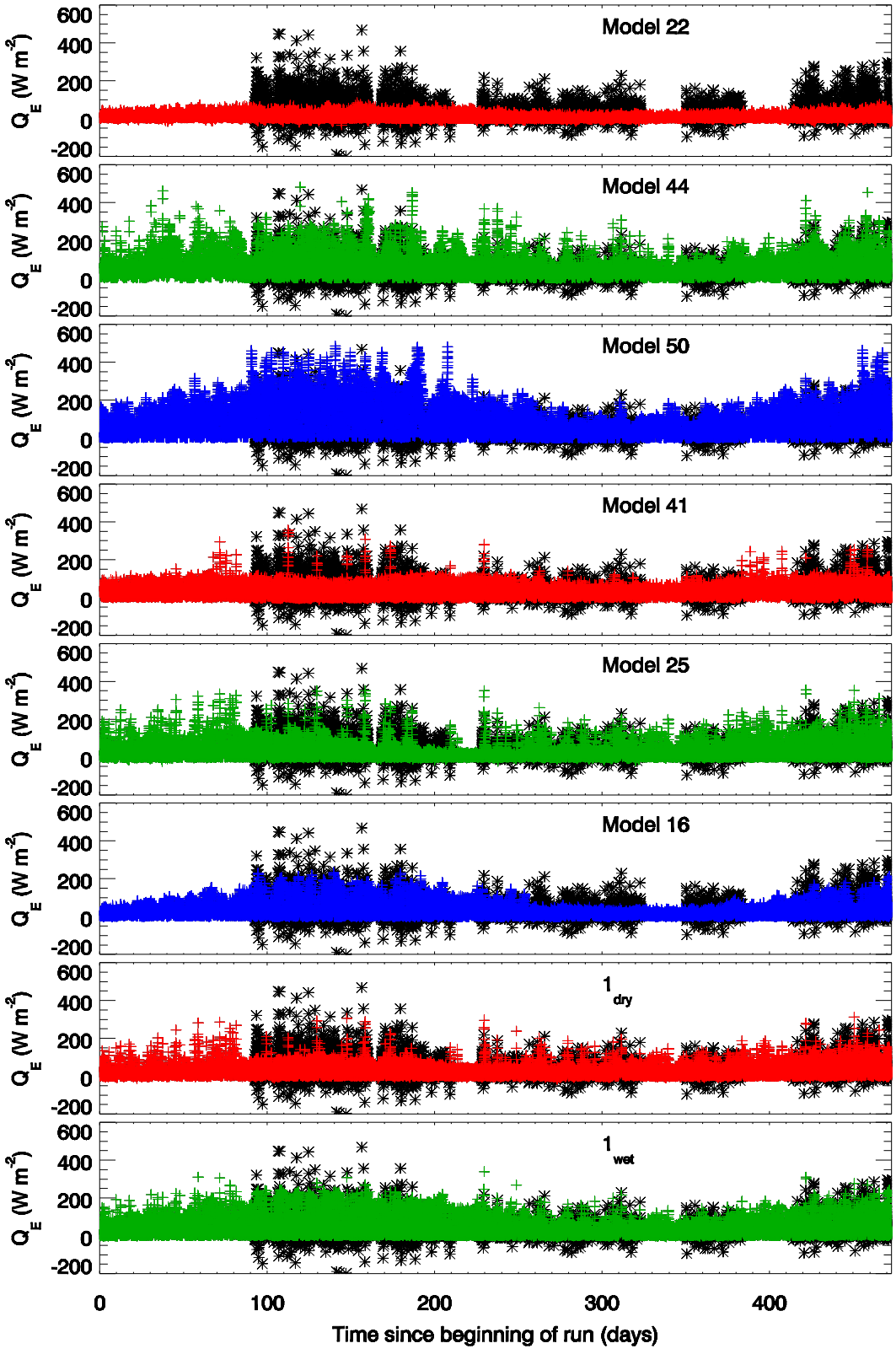

Figure 5.2: Time series of observed and modelled latent heat flux for a selection of models (Black asterisk – observations, coloured pluses – model)

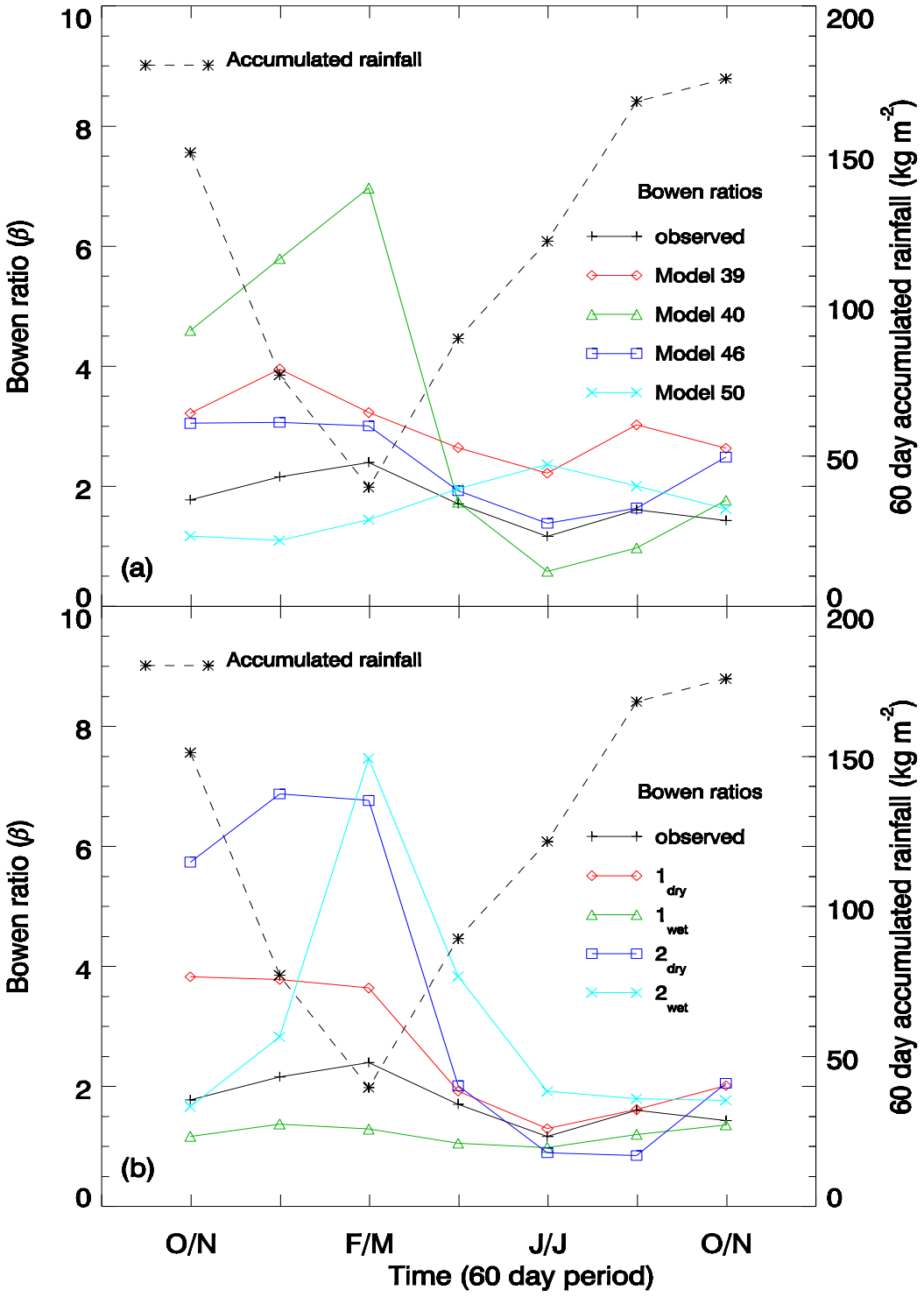

Figure 5.3: Evolution over the seasonal cycle of the Bowen ratios calculated from the mean midday (10:00 – 14:00) fluxes for (a) a selection of models (see text) (b) two urban LSM ("1", "2") each run by two separate groups.

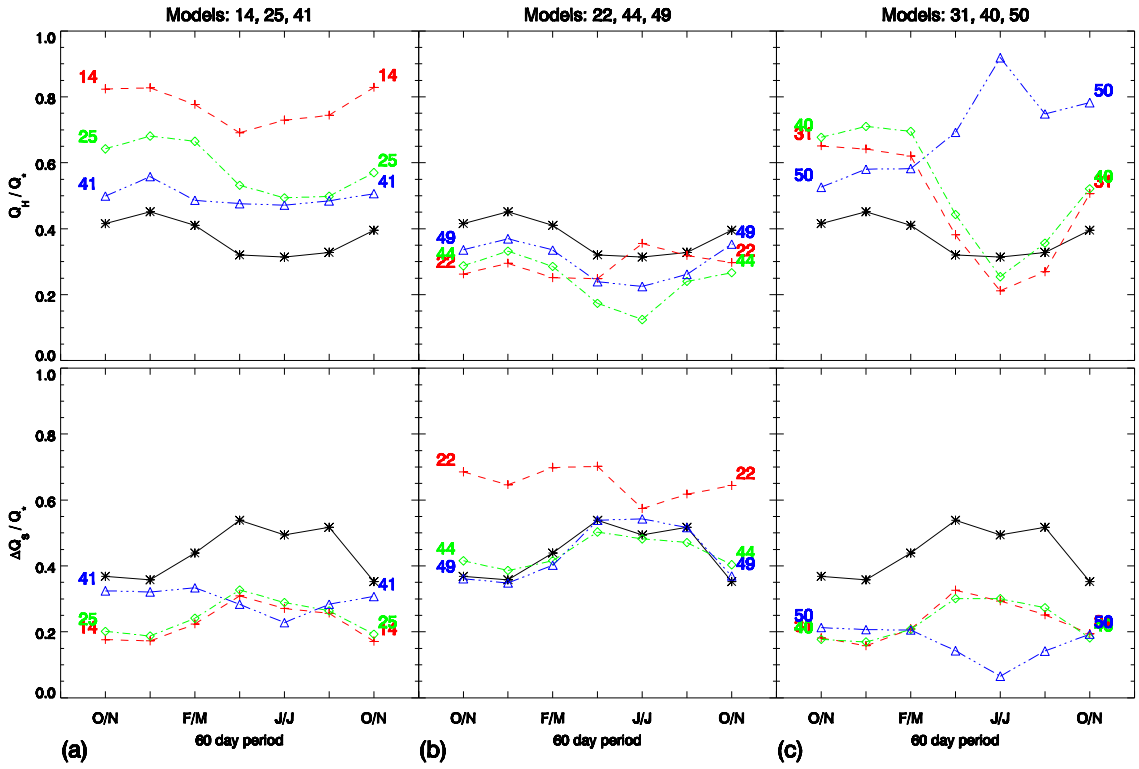

Figure 5.4: Observed (black line *) and modelled ratio of the average midday (10:00 – 14:00) (top) turbulent sensible heat flux to the net all wave radiation flux (Q_H /Q*) and(bottom) net storage heat flux to the net radiation flux (ΔQ_S /Q*) over the seasonal cycle for models with: (a) typical performance, (b) Q_H /Q* smaller than observed, and (c) Q_H /Q* smaller than observed in winter and spring along with model 50.

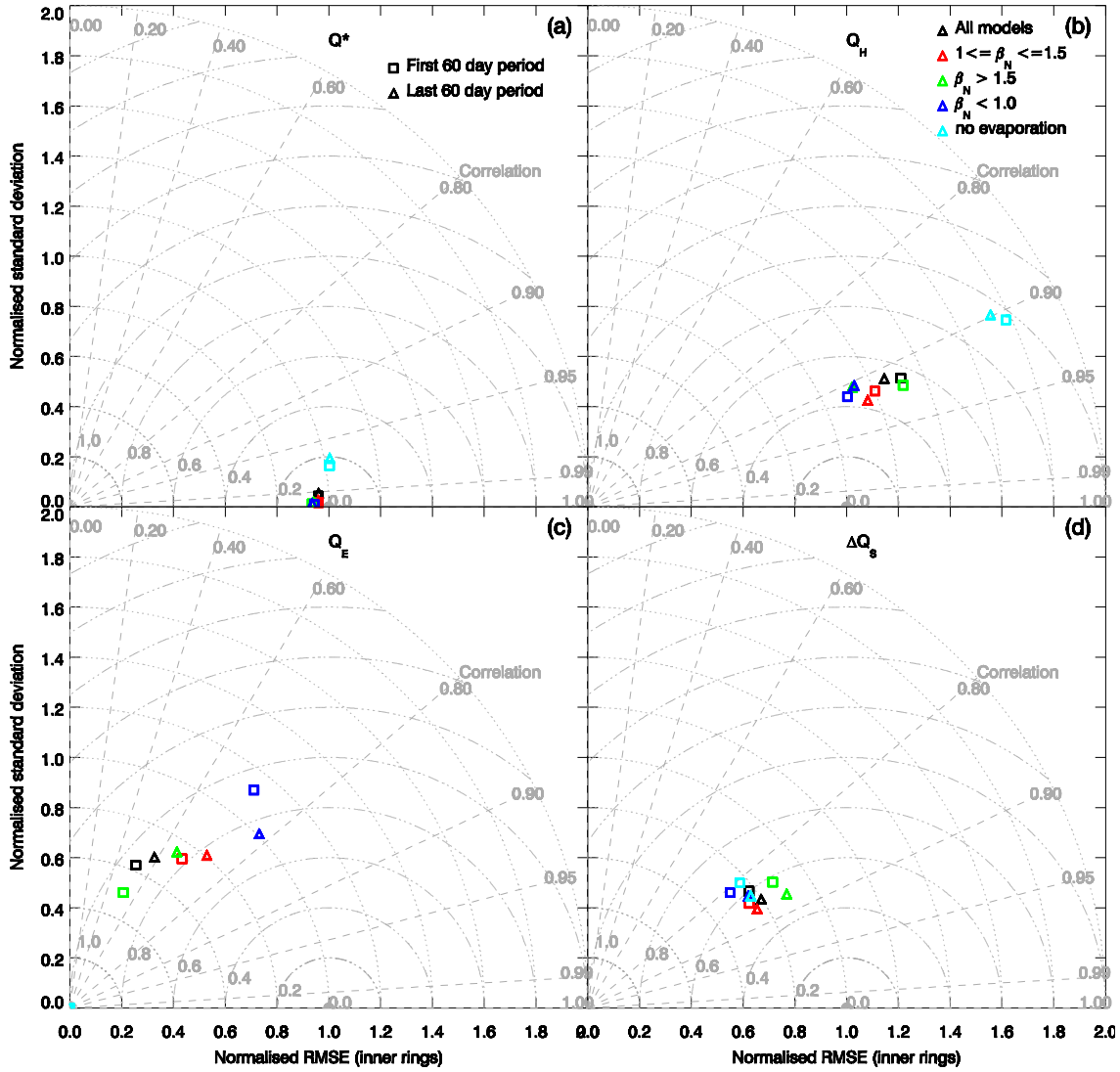

Figure 5.5: Normalised Taylor plot for first (square) and last (triangle) 60 day periods for (a) net all wave radiation, (b) turbulent sensible and (c) latent heat and (d) net storage heat fluxes. Taylor plots have the correlation coefficient on the polar axis, the normalised standard deviation on the radial y-axis and the normalised *RMSE* (x-axis) on the internal circular axes (Taylor 2001). Performance for the average of all the models, plus the average of the models grouped by their normalised Bowen ratio (β_N), as identified in Fig. 5.2 (see key). For (c) models with no evaporation the points plot at the origin.

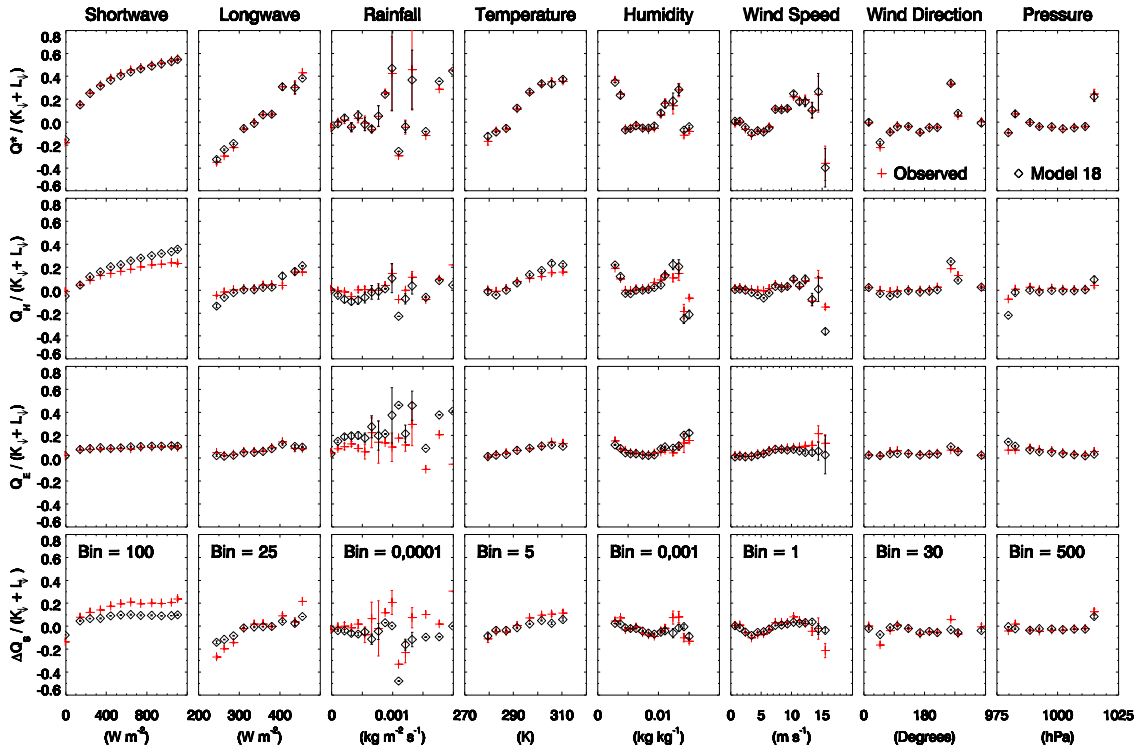

Figure 5.6: Observed and a typical model (model 18) fluxes normalized by observed incoming all wave radiation ($Q\downarrow = K\downarrow + L\downarrow$) binned (bin sizes as indicated on each figure, units as per figure) by atmospheric forcing type: median (symbol) and standard error (vertical bar)

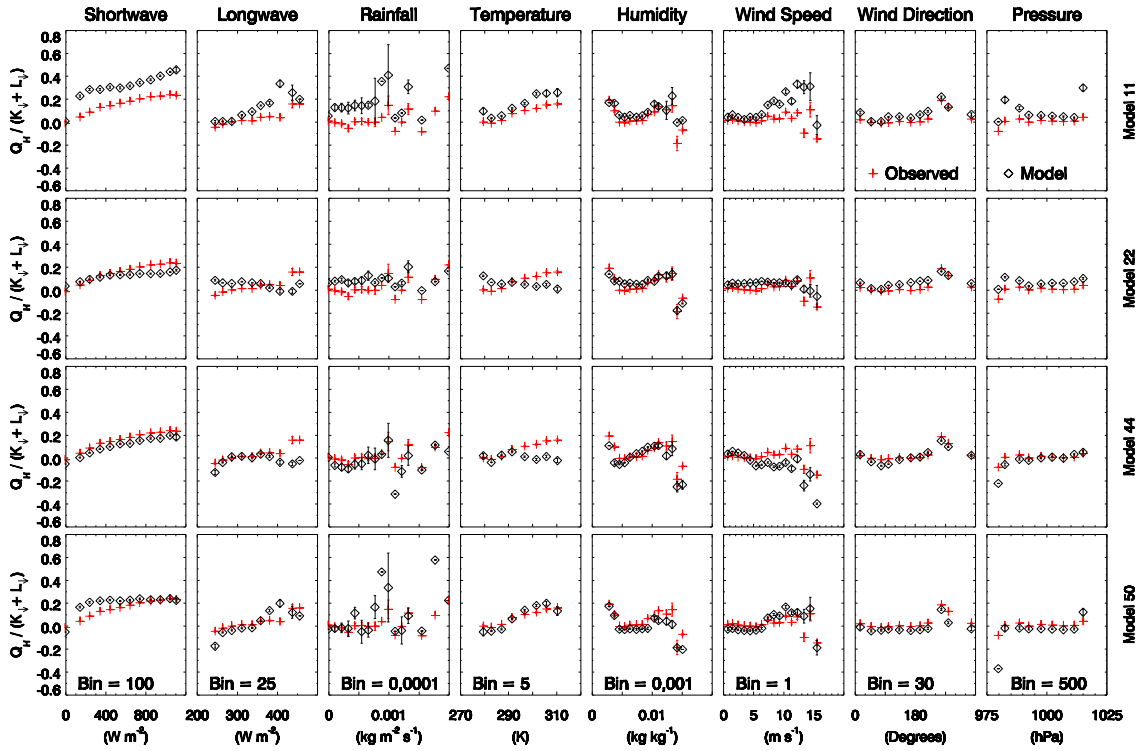

Figure 5.7: Sensible heat flux normalized by incoming all wave radiation ($Q\downarrow = K\downarrow + L\downarrow$) binned (bin sizes as indicated on each figure, units as per figure) by atmospheric forcing type (columns) for four models (11, 22, 44, 50 rows) with characteristics that differ from the observations: median (symbol) and standard error (vertical bar)

Chapter 6: Key conclusions of the first international urban land surface model comparison project[12]

Abstract

The first international urban land surface model comparison was designed to identify three aspects of the urban surface-atmosphere interactions: (1) the dominant physical processes, (2) the level of complexity required to model these, and (3) the parameter requirements for such a model. Offline simulations from 32 land surface schemes, with varying complexity, contributed to the comparison. Model results were analysed within a framework of physical classifications and over four stages. The results show that the following are important urban processes; (i) multiple reflections of shortwave radiation within street canyons, (ii) reduction in the amount of visible sky from within the canyon, which impacts on the net long-wave radiation, (iii) the contrast in surface temperatures between building roofs and street canyons, and (iv) evaporation from vegetation. Models that use an appropriate bulk albedo based on multiple solar reflections, represent building roof surfaces separately from street canyons and include a representation of vegetation demonstrate more skill, but require parameter information on the albedo, height of the buildings relative to the width of the streets (height to width ratio), the fraction of building roofs compared to street canyons from a plan view (plan area fraction) and the fraction of the surface that is vegetated. These results, whilst based on a single site and less than 18 months of data, have implications for the future design of urban land surface models, the data that need to be measured in urban observational campaigns, and what needs to be included in initiatives for regional and global parameter databases.

6.1. Introduction

Urban areas are often warmer than their surrounding rural environments, referred to as the urban heat island (UHI). This urban warming has numerous effects, including the initiation of convective storms (e.g., *Bornstein and Lin*, 2000), altering pollution dispersion by adapting mixing through changes to atmospheric boundary layer structure (e.g., *Sarrat et al.*, 2006, *Luhar et al.*, 2014), impacts on the production and mixing of ozone (e.g., *Chaxel and Chollet*, 2009,

[12] *This has been published as:* Best MJ, Grimmond CSB (2015) Key conclusions of the first international urban land surface model comparison. Bulletin of the American Meteorol. Soc., 96, 805-819. doi:http://dx.doi.org/10.1175/BAMS-D-14-00122.1

Ryu et al., 2013), enhanced energy demand for summer-time cooling through air conditioning (e.g., *Radhi and Sharples*, 2013, *Li et al.*, 2014), impacts on urban ecology (e.g., *Pickett et al.*, 2008, *Francis and Chadwick*, 2013) and increased mortality rates during heat waves (e.g., *Laaidi et al.*, 2012, *Herbst et al.*, 2014, *Saha et al.*, 2014). As such, it is important to be able to accurately forecast urban warming and other meteorological variables for cities where the majority of the World's population now lives.

Predictions of future climate suggest additional warming in urban environments (*McCarthy et al.*, 2010, *Oleson et al.*, 2011). Indeed, the Inter-Governmental Panel on Climate Change (IPCC) Working Group 1 Fifth Assessment Report (*IPCC*, 2013) included at least one model that explicitly included an urban representation, and this number is likely to increase in the future as the resolution of these climate models increases to the extent that some urban areas are resolved. For future design of buildings and planning of cities, it is important that the dominant processes that lead to urban warming effects are considered. This requires the development of models that can represent the most important features of the urban heat island be used for reliable predictions.

The urban heat island results from differences in surface energy exchanges between the urban environment and its surrounding rural area. Thus, understanding these differences is needed to interpret the urban heat island. The differences in urban surface energy exchanges arise through a number of processes. The geometry of a street canyon will increase the incoming solar radiation and long-wave radiation that are absorbed, due to multiple reflections and re-radiated from the 3-dimensional structures. The orientation of street canyons and the elevation of the sun will impact the reflected solar radiation, as a consequence of the depth to which the direct sunshine can penetrate into the canyon. The reduced availability of water at the urban surface, compared to natural vegetated or bare soil surfaces, means more of the incoming solar radiation is transformed into heat rather than a flux of moisture into the atmosphere. However, a larger proportion of this energy for heating is held within the fabric of the buildings given the large thermal inertia of the materials, resulting in changes in the diurnal cycle of urban temperatures. Moreover, an additional source of heating within the urban areas comes from human activities such as transport, the internal heating of the buildings and the metabolic rates of the people themselves (e.g., Sailor and Lu, 2004).

All of these processes contribute to the differences in the energy balance between urban and rural surfaces, but it is difficult to identify which are the dominant processes just from observations as the processes cannot be separated because of the complex nature of the environment. As such, the best way to study these processes individually is by using urban land

surface models (ULSMs) that have been developed for weather and climate applications, i.e., exchange surface fluxes with an atmospheric model. There are a number of such ULSMs that vary considerably in their complexity (e.g., *Kusaka et al.*, 2001, *Fortuniak*, 2003, *Krayenhoff and Voogt*, 2007, *Hamdi and Masson*, 2008, *Lee and Park*, 2008, *Oleson et al.*, 2008a). Although newer models often include more complex features than previous models, without knowing the dominant processes and controls, it is difficult to quantify the impact of each new feature.

The first urban land surface model comparison was designed to objectively assess and compare the performance of a range of ULSMs for a single observational site. It attempted to identify the dominant physical processes that need to be represented in ULSMs by comparing models of varying complexity (Table 6.1). These models ranged from simple bulk representations of the surface that have been applied to atmospheric models for over a decade, representations of the facets of a street canyon (i.e., roofs, walls and road) that have been used in weather and climate models, through to more recently developed schemes that consider a complete energy balance at various levels within the urban canyon that have been applied to stand alone single point studies. Figure 6.1 shows a conceptual representation of the surface energy balance for these models of varying complexity. Whilst the scale that these models typically represent is larger than the size of the elements within a street canyon, a common feature is the ability to predict the exchange of fluxes between the urban surface and the atmosphere above it, i.e., the net all-wave radiation (Q^*), turbulent sensible (Q_H) and latent heat (Q_E) fluxes, as measured from flux towers in numerous urban observational campaigns.
The aim of the urban model comparison was to consider:
(1) What are the dominant physical processes in the urban environment?
(2) What is the level of complexity required for an ULSM to be fit for purpose?
(3) What are the parameter requirements for such a model?
Here we present an analysis of the model comparison results to address these questions.

6.2 Model Comparison design

6.2.1 Observational data

The criteria for selecting the evaluation dataset were; first it had not been used to evaluate any ULSMs previously, and second it needed to cover an annual cycle to allow assessment for different seasons. Model evaluation studies often result in the development and optimisation of a model in order to obtain better representation of the assessed metrics. Hence, using a dataset

previously used by one or a sub-set of the models to be evaluated would not enable a clean/independent objective assessment for all of the models.

The dataset for a suburb of Melbourne (Preston) (*Coutts et al.*, 2007a, 2007b) that had observations from 13 August 2003 to 13 November 2004 was selected. The moderately developed, low-density housing area is classified by *Coutts et al.* (2007b) as an Urban Climate Zone (UCZ) 5 (*Oke*, 2006), Local Climate Zone (LCZ) 6 (*Stewart and Oke*, 2012) or *Loridan and Grimmond* (2012b) Urban Zone for Energy exchange (UZE) medium density. The description of UCZ 5 is "medium development, low density suburban with 1 or 2 storey houses, e.g., suburban housing" (*Oke*, 2006), and as such the site is typical of suburban areas found in North America, Europe and Australasia. The area has mean building height-to-width ratio of 0.42 and mean wall-to-plan ratio of 0.4 (*Coutts et al.*, 2007b). The surface is dominated by impervious cover (44.5% buildings, 4.5% concrete and 13% roads), with a pervious cover of 38% (15% grass, 22.5% other vegetation and 0.5% bare ground or pools) (*Coutts et al.*, 2007a).

The methods used to obtain the observed fluxes applied to our current analysis are given in Table 6.2, with details (e.g., data processing) presented in the original observation papers (*Coutts et al.*, 2007a, 2007b). In addition, the initial model comparison results papers (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013, 2014) provide the site parameters. A continuous gap-filled atmospheric forcing dataset (474 days) to run the models was created for this study (see *Grimmond et al.*, 2011). To evaluate the modelled fluxes (sensible heat flux, latent heat flux, net all-wave radiative flux and net storage heat flux (ΔQ_S)) 30 min periods are used when no observed fluxes are missing to allow consistent analysis between the fluxes (N=8865 or 38.9% of the full period).

6.2.2 Data analysis

To permit the research questions posed above to be considered, information about the observational site was released to the modelling groups in stages. This enabled analysis of the importance of the different types of information to model performance through assessment of the change in model skill between the stages. The stages (Table 6.3), designed to correlate with ease of access to information for all cities globally, involved release of (*Grimmond et al.*, 2011):
Stage 1: Atmospheric forcing data: (Table 6.3), typically provided by an atmospheric model.
Stage 2: Vegetation and built fraction: two dimensional plan area characteristics of the site. These can be determined from land cover datasets derived from satellite data.
Stage 3: Morphology: three dimensional characteristics of the site (Table 6.3.). These can be

interpreted from LiDAR (e.g., *Goodwin et al.*, 2009, *Lindberg and Grimmond*, 2011), aerial photographs (e.g., *Ellefsen*, 1990/1991), detailed satellite imagery (e.g., *Brunner et al.*, 2010), or simple empirical relations (e.g., *Bohnenstengel et al.*, 2011).
Stage 4: Building material parameters (Table 6.3): only obtainable from local knowledge of the materials used in the construction of the buildings.
Stage 5: Observed fluxes: to allow parameter optimisation studies. Only a few groups completed this stage, so these results are not presented here.

The results from 24 modelling groups are analysed, involving 21 independent models (Table 6.1). Alternative versions of the same model were run by the same or independent modelling groups, which resulted in 32 sets of model simulations being submitted for all of the four stages (see full list in *Grimmond et al.*, 2011). Each group completed a survey indicating the level of complexity used for various physical processes within their models. From the latter, categories of physical processes were established, with classes that cover the range of complexities (*Grimmond et al.*, 2010, 2011). These categories were chosen to investigate the importance of various physical processes that could contribute to differences in the surface energy balance between the urban and rural environments. Thus every model is assigned to a class in each category based on the survey information. In this study, the complexity category (Grimmond et al., 2011) is not considered as the focus is to separate the specific physical processes. The categories, with the number of models in each class are shown in Table 6.4.

Comparing the mean behaviour of the models in each of the classes as a reference provides a method to determine the level of complexity that gives the best performance for each category. These data are analysed to address the second research question, where "fit for purpose" in this study is defined as being able to accurately represent the energy exchange between the urban surface and the atmosphere (i.e., the net all-wave radiation, turbulent sensible and latent heat fluxes).

Furthermore, by assessing the performance of the models across the categories for all classes, it is possible to identify the physical processes that have the largest impact on the performance of the models, hence identifying the dominant physical processes and addressing the first research question.

6.2.3 Methodology

Initial results from the urban model comparison (*Grimmond et al.*, 2011) ranked the models and

assessed the performance of the various classes within the categories using standard statistical measures. Here an alternative approach to assess the models' performance is used, that considers the percentage of the models' data values that are within observational error (E_{obs}). This gives a measure between zero (no values within observational errors) and 100% (all values within observational errors, i.e., a 'perfect' model). Although this type of analysis is not strictly benchmarking, as each model is not being compared to an *a priori* metric, it could be considered as being closer to the benchmarking ethos as having all data points within observational errors would be a stringent metric.

The observational error estimates used in this analysis are for day-time fluxes based on a percentage of the observed fluxes, as suggested by *Hollinger and Richardson* (2005): net all-wave radiation flux 5%, turbulent sensible heat flux 10%, latent heat flux 8%, and upward components of both shortwave and long-wave radiation fluxes 10%. As the net storage heat flux in the observational dataset is determined as the residual of the surface energy balance, its observational error is assumed to be the sum of the errors for the other terms (i.e., Q^*, Q_H and Q_E), giving 23%. The night-time error estimates are assumed to be double the day-time error estimates for each of the fluxes. The absolute magnitude of fluxes during this period are typically small (order of (10) W m^{-2}), hence changes in the percentage of the observed flux used as the error estimates are likely to be within the reporting resolution (e.g. order of (1) W m^{-2}) of the observations (especially the turbulent fluxes). Whilst these error estimates may be indicative rather than the actual values, the results would not substantially change the analysis presented.

The analysis was undertaken for each model (k) in each class (j) within each category (i) (Table 6.4), for each flux, over each stage within the comparison, and separately for day-time and night-time. From this the percentage of data within observational error ($E_{obs\ i,j}$) was determined:

$$E_{obs,ij} = \frac{\sum_{k=1}^{n_{ij}} M_k}{n_{ij} T} \times 100\% \qquad (6.1)$$

where M is the number of points within observational error for model (k), n is the number of models and T is the number of day-time or night-time points in the time series as appropriate.

6.3. Results

Application of eqn. 6.1 to the sensible, latent and net storage heat fluxes, for each class and category, at Stage 1 and Stage 4 (Table 6.3) are shown in Figure 6.2. The results could range between 0% (i.e., no model data points within the observations errors) to 100% (i.e., all model

data points within observational errors). The relative changes between the stages are also shown if Figure 6.2, i.e., for stage (s) the change relative to the previous stage (s-1) given by:

$$E^{s}_{obs,ij} / E^{s-1}_{obs,ij} \quad (6.2)$$

Assessment of "between stages performance" allows an emphasis of the common results across all of the classes and categories. It is scaled between 0% and 100%, with 50% corresponding to no change between the stages (Figure 6.2).

Generally the results of the analysis, consistent with *Grimmond et al.* (2011), show that the skill to model latent heat fluxes is improved between stages 1 and 2. Knowing the plan area vegetation fraction (provided in Stage 2) is important for modelling the latent heat flux. No other stages show a general increase in model performance across the classes and categories for the fluxes shown in Figure 6.2. For the radiation fluxes (Fig. 6.3), the largest changes evident between Stages 3 and 4 are for the reflected shortwave radiation flux and are due to the specification of the bulk albedo at the site (i.e., the ratio of the reflected outgoing shortwave radiation flux from the whole urban surface to the incoming shortwave radiation flux, information released at Stage 4). This is also consistent with the conclusions from *Grimmond et al.* (2011).

Model performance for the outgoing long-wave radiation flux has its largest changes at night-time between Stages 2 and 3 (when the 3-d site morphological information (Table 6.3) were made available, Fig. 6.3). This enhanced performance at night could be related to improved estimates of the sky view factor which influences radiative trapping, and/or from improved estimates of the difference in nocturnal surface temperatures between building roofs and those of the roads and walls of the urban canyons. Improved performance is not detected in the day-time outgoing long-wave radiation flux (Fig. 6.3), probably because of the dominance of shortwave radiation at this time. These results were not identified in *Grimmond et al.* (2011) as there was no separate analysis for day-time and night-time.

It is evident from Figures 6.2 and 6.3 that the performance of the models for each of the fluxes does not improve consistently for each stage, as might be expected. This suggests that the models are not able to correctly make use of all of the information that is provided at each of the stages and hence the design of the models, and the use of their specific parameters, is not necessarily correct. This is discussed further in Grimmond et al. (2011).

Each model is assigned to one class for every category (Table 6.4). This means that a model with particularly good (or poor) performance will influence the results for its class in each of the

categories. The implications of this are that it is not possible to ensure that the good performance from a particular class within one category is not actually resulting from the results of a class from a different category. This potential contamination of results by categories inhibits the analysis of the dominant physical processes and the suitability of the models. Both the analysis presented in *Grimmond et al.* (2011) and that in Figures 6.2 and 6.3 have this limitation, hence we will not consider further any results in Figures 6.2 and 6.3 for any specific class or category. Alternatively, to address this issue of cross-contamination, we repeat the complete analysis using eqn. 6.1 separately for each category (c), but only considering the subset of models from class (*a*). Hence for each class (*j*) in category (*i*) for the analysis of eqn. 6.1, the models used are those that are in both class (*a*) of category (*c*) and class (*j*) of category (*i*), of which there are $n_\alpha = n_{ca} \bigcap n_{ij}$, thus:

$$E_{obs,caij} = \frac{\sum_{k=1}^{n_\alpha} M_k}{n_\alpha T} \times 100\% \tag{6.3}$$

This gives the equivalent of 26 versions of Figures 6.2 and 6.3 (one for each class in each category); although for a given subset of models it is inevitable that some classes will not have any members and hence have no data. We then apply the following equation for each of the stages to determine which of the original class of models has the best performance:

$$P_{ca} = \frac{\sum N_m}{N_{tot} - \left(\sum N_{nd}\right) - 1} \times 100\% \tag{6.4}$$

where P_{ca} is the percentage of classes in the analysis that are improved from just the subset of models (compared to the analysis with the full set of models),

$$N_m = \sum_{k=1}^{N_{tot}} \begin{cases} 1 \text{ if } E_{obs,caij} > E_{obs,ij} \\ 0 \text{ otherwise} \end{cases} \tag{6.5}$$

is the number of classes that are improved in the analysis, N_{tot} is the total number of classes ($\sum ij = 26$) and

$$N_{nd} = \sum_{k=1}^{N_{tot}} \begin{cases} 1 \text{ if } n_{ca} \bigcap n_{ij} = 0 \\ 0 \text{ otherwise} \end{cases} \tag{6.6}$$

is the number of classes with no data.

Hence values of P_{ca} close to 100% relate to nearly all classes in all categories being improved from the physical process represented in class (a) of category (c). This indicates that this process and its representation are important to model performance. Whereas values close to 0% relate to almost all classes in all categories being degraded, suggesting that the representation of the

physical process is detrimental to model performance. Values around 50% have a similar number of classes that are improved and degraded, suggesting that the representation of the physical process has little impact on model performance. Hence the conclusions that can be drawn from this analysis are more robust than those of Figures 6.2 and 6.3, and the previous study of *Grimmond et al.* (2011).

For example, with models that have an infinite number of reflections (category R, class i), the median of the results over the stages give a value of 88% for the night-time net storage heat flux (Fig. 6.4). This results from 14 of the 16 possible classes containing data being improved when considering only these models, demonstrating that this is important for predicting this flux. However, models that have multiple reflections (category R, class m) have a value of 12.5% for the night-time net storage heat flux (Fig. 6.4). This results from only two of the possible 16 classes containing data being improved, hence showing that this is detrimental to predicting the flux.

The results of Figure 6.4 show that for some classes (e.g., infinite reflections; category R, class I, Table 6.4), there are some demonstrated improvements to a flux (e.g., LW_{up}) which is not obviously explained by the physics (e.g., how do infinite reflections of shortwave radiation improve the outgoing long-wave radiation but not the reflected shortwave?). Also, there are some classes that improve one particular flux, but not other fluxes. For example, models that represent the net storage heat flux as the residual of the surface energy balance (category S, class r, Table 6.4) demonstrate a clear improvement for the day-time sensible heat flux, but not for the latent or the net storage heat fluxes. This could be because with such models the sensible heat flux is not constrained by the energy balance giving them the freedom to enable better predictions of the sensible heat flux, whilst moisture availability is still the main control for the latent heat flux.

There are many such conclusions that can be drawn from Figure 6.4. Here the focus is on results that are consistent between the fluxes, or consistent for a particular flux between the day-time and night-time.

Models with a bulk representation of the albedo and emissivity (category A_E, class 1, Table 6.4), and a bulk representation of facets and orientation (category F_O, class 1; the models in these two classes were identical), demonstrate an improvement in skill during the day-time for nearly all fluxes, with the exceptions of the outgoing long-wave radiation which shows little change in skill and net all-wave radiation fluxes with only small improvements (Fig. 6.4). This class of models also shows an improvement in the night-time sensible and latent heat fluxes, but

degradation in the radiative fluxes during the night. These improved results are most likely due to the ability to utilize the observed bulk albedo directly. This class of models clearly delivers the largest benefits across the fluxes and indicates the most significant physical process to represent is the bulk albedo for the urban surface, because the net shortwave radiation dominates the surface energy balance.

Improvements to the outgoing long-wave radiation flux and the net all-wave radiation flux during both day-time and night-time are obtained from models that have a single layer for each element of the urban environment (i.e., roofs and either urban canyons, or walls and roads separately) in the morphology category (category L, class 2, Table 6.4; Fig. 6.4). Improvements to the night-time sensible heat flux and net storage heat flux are also obtained from this class of models, but there is no improvement to these fluxes during the day-time. This neutral day-time result in the sensible and net storage heat fluxes may be explained by the negative impact on the outgoing shortwave radiation flux, which dominates over the long-wave radiation flux during the day-time. However, these results demonstrate the importance of presenting the difference in radiative surface temperatures between the roofs and the urban canyon, due to the non-linear relationship between the upward long-wave radiation and the radiative temperature.

When considering the way in which the models represent vegetation (category V, Table 6.4), we find that although including vegetation (classes s and i, Table 6.4) does generally lead to an improvement for the fluxes, these improvements are not as obvious as those from the bulk albedo or the single layer urban morphology. Hence although these results confirm those presented in earlier studies on the comparison (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013, 2014), that representing vegetation gives improved results, we demonstrate that the more robust analysis presented here shows that this is not the most important physical process as was concluded in these earlier studies. Getting the radiative fluxes correct from the shortwave via the bulk albedo and the long-wave through the urban morphology are required before the vegetation can influence the partitioning of energy between the sensible and latent heat fluxes.

Previous studies on the urban comparison data have also concluded that models which neglect the anthropogenic heat flux (Q_F) do at least as well as the models that include this flux, although they were unable to explain this result (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013, 2014). However, the results in Figure 6.4 show that although the class of models that neglect the anthropogenic heat flux (category A_N, class n, Table 6.4) do improve some of the fluxes, the improvements are not consistent over all of the fluxes. Moreover, this class of models within the anthropogenic heat flux category is not always the one that delivers the best results. Hence we can conclude that although the models that neglect the anthropogenic heat flux do show some

improved results, we cannot make any significant statements about the classes within this category.

6.4. Conclusions

Prior conclusions from the ULSM comparison with daily (24 h) and seasonal analysis include that: representation of vegetation is critical to model performance (*Grimmond et al.*, 2011, *Best and Grimmond*, 2013), along with the associated initial soil moisture (*Best and Grimmond*, 2014), and the bulk albedo is also important (Grimmond et al., 2011). Notably, neglecting the distinctive urban anthropogenic heat flux was not found to penalize performance (albeit in the suburban area the value is small) (*Best and Grimmond*, 2013). However, this new analysis considering diurnal performance (day, night) enables us to conclude that nocturnal radiative processes also benefit from accounting for the enhanced long-wave trapping that occurs within urban areas. Separating the radiative processes of the roof and the urban canyon is beneficial.

More critically, the more robust analysis presented here enables identification of a re-prioritisation of the key physical processes: firstly, ensuring the use of the correct bulk albedo for the urban surface; secondly, the outgoing long-wave radiative fluxes with the representation of morphology separated into roofs and urban canyons; and thirdly, the inclusion of vegetation. The implications of the bulk albedo is important for observations as the temporal resolution of satellite estimates mean they will not provide the variations by time of day that are observed (e.g., Christen and Voogt, 2004, Grimmond et al. 2004a, Kotthaus and Grimmond 2014b).

The current results for anthropogenic heat flux are consistent with the earlier studies: that neglect of the relatively small magnitude flux at this site (study period mean = ~17 W m^{-2}) is reasonable. This conclusion could well be different for urban environments where this is a more significant term in the surface energy balance. The flux is expected to be larger in other areas of Melbourne (e.g., as suggested from analysis using the model of Lindberg et al. 2013) and for urban areas elsewhere. We therefore recommend that future model comparisons ideally include areas of cities with larger anthropogenic heat fluxes.

Thus to answer the three over-arching research questions of the urban model comparison: The dominant physical processes in the urban environment that models need to be able to simulate, in order, are; changes to the bulk albedo of the surface that result from building materials and also shortwave trapping from the canyon geometry; the reduction in outgoing long-wave radiation from the street canyon due to a reduced sky view factor and the contrast

between this and the roofs that see a full sky view; and the evaporation from vegetation.
For the current generation of ULSMs, the ability to utilize a bulk surface albedo (category A_E, class 1, Table 6.4) and to be able to distinguish between the roofs of buildings and the urban canyons (category L, class 2), and to have a representation of vegetation (category V, classes s, i), results in the best performance.
The key parameters for ULSMs are the bulk surface albedo (information given for Stage 4 influencing the upward shortwave radiation flux), the height to width ratio of the urban canyons and the fraction of building roofs to the urban canyons (information given for Stage 3 influencing the upward long-wave radiation flux), and the vegetation fraction (information given for Stage 2 influencing the sensible and latent heat fluxes).

The results, from this and the previous studies on the ULSM comparison, all suggest that a simple representation for most of the physical categories is sufficient for this type of application, i.e., determination of local scale fluxes (e.g. for use in the coupling to an atmospheric model). The prior categorization of the models (Grimmond et al., 2011, Best and Grimmond, 2013) into (simple, medium and complex) complexity classes based upon the number of physical categories treated as simple by a model demonstrated that the simple models performed best. This relative success of simple models suggests that for simulating local scale fluxes, more complex schemes deliver little additional benefit. Furthermore, the reduced parameter requirements for simple schemes are advantageous for large scale applications, such as global or regional scale modelling. However, it cannot be expected that this conclusion would also hold for other applications, e.g., atmospheric dispersion within street canyons of a specific city, as the simple models do not present some of the basic physical requirements for such applications. Thus the requirement for the development of more complex ULSMs does remain.

The implications of this study go beyond the urban environment. In general, we need to balance the requirement for complexity within models against what is actually required for a model to be fit for purpose. Hence new and more complex processes should not be included in models unless it can be demonstrated that they are required. In addition, consideration needs to be given to the availability of information to specify parameters within complex models, and if such complexity can be justified given the uncertainty range for the parameters. Also, the type of analysis used here could be applied to any comparison study to ensure that the results are robust and not contaminated by physical processes not being directly considered.

These key conclusions are based on the single site observational dataset of less than 18 months. This suburban site of low density housing, is typical of extensive areas in North America,

Europe and Australasia. Hence we might expect the results from this study to be valid over a reasonable range of cities. However, most urban environments have a range of zones (e.g. *Ellefsen*, 1991, *Grimmond and Souch*, 1994, *Stewart and Oke*, 2012) with very different characteristics. So to test if the results presented here are robust for other cities, similar "experiments" are required for additional sites with differing climates and urban characteristics. Hence we recommend that further model comparison projects are required for the urban community.

Despite these limitations, the results have implications for future development of ULSMs and for the types of data that need to be collected in future urban measurement campaigns (e.g., soil moisture, given its impact to limit transpiration and the long timescales required for model spin-up, along with the conclusion that the fraction of vegetation is important for urban areas) and/or the parameters that should be collated systematically for cities around the world (e.g., *Ching et al.*, 2009, *Loridan and Grimmond*, 2012b, *Stewart and Oke*, 2012, *Ching*, 2013, *Faroux et al.*, 2013).

Table 6.1: Urban land surface models (ULSMs) used to obtain results that are analysed here. See Grimmond et al. (2010, 2011) for more details of the different model versions and the number of groups that submitted simulations to the urban model comparison.

Model name	References
Building effect parameterization (BEP)	Martilli et al. (2002) Salamanca et al. (2009, 2010) ; Salamanca and Martilli (2010)
Community Land Model – urban (CLM-urban)	Oleson et al. (2008a, 2008b)
Institute of Industrial Science urban canopy model	Kawamoto and Ooka (2006, 2009a, 2009b)
Joint UK land environment simulator (JULES)	Essery et al. (2003); Best (2005); Best et al. (2006); Best et al. (2011)
Local-scale urban meteorological parameterization scheme (LUMPS)	Grimmond and Oke (2002); Offerle et al. (2003); Loridan et al. (2011)
Met Office Reading urban surface exchange scheme (MORUSES)	Harman et al. (2004a, 2004b); Harman and Belcher (2006), Porson et al. (2010)
Multi-layer urban canopy model	Kondo and Liu (1998); Kondo et al. (2005)
National and Kapodistrian University of Athens model	Dandou et al. (2005)
Noah land surface model/single-layer urban canopy model	Kusaka et al. (2001); Chen et al. (2004); Loridan et al. (2010)
Seoul National University urban canopy model	Ryu et al. (2011)
Simple urban energy balance model for mesoscale simulation	Kanda et al. (2005a, 2005b); Kawai et al. (2007, 2009)
Slab urban energy balance model	Fortuniak (2003); Fortuniak et al. (2004, 2005)
Soil model for submesoscales (urbanized)	Duport and Mestayer (2006); Dupont et al. (2006)
Temperatures of urban facets (TUF)	Krayenhoff and Voogt (2007)
Town energy balance (TEB)	Masson (2000); Masson et al. (2002); Lemonsu et al. (2004); Pigeon et al. (2008), Hamdi and Masson (2008)
Vegetated urban canopy model	Lee and Park (2008)

Table 6.2: Methods used to obtain the observed fluxes used for comparison with the ULSM. Sources: *Coutts et al*., (2007a, 2007b). Height of observation for all fluxes: 40 m.

Flux	Instrument / Method	Sampling frequency (Hz.)	Averaging period (min)
SW_{up} LW_{up} Q^*	Kipp and Zonen CM7B and CG4 radiometers	1	30
Q_H	CSI CSAT 3D sonic anemometer	10	30
Q_E	CSI CSAT 3D sonic anemometer CSI Krypton hygrometer (Aug 2003 – Feb 2004), LiCOR LI7500 open-path infrared gas analyser (remaining period)	10	30
ΔQ_S	Residual of the surface energy balance	N/A	30
Q_F	Calculated (*Sailor and Lu* ,2004) : *Vehicles*: Numbers from survey (Nov. 2002 – Oct 2003 *Building* sector: 30 min electricity and daily natural gas statistics *Human metabolism*: Night, day and transition period metabolic rates, with population density statistics	N/A	Average monthly diurnal cycle at 30 min. resolution

Table 6.3: Information released at each stage of the comparison

Stage	Information released
1	Atmospheric forcing data only (incoming shortwave radiation, incoming long-wave radiation, precipitation, atmospheric wind speed, temperature, specific humidity and surface pressure)
2	Vegetation and built fractions
3	Morphology (Building heights, height-to-width ratio, mean wall to plan area ratio, fraction of surface covered by buildings, concrete, road,)
4	Specific information on building materials (e.g., albedo and thermal properties of wall, road, roof)
5	Observed fluxes for parameter optimisation (Not considered in this study)

Table 6.4: Classes and physical categories used in the analysis of the urban comparison results, including the number of models in each class (see *Grimmond et al.*, 2010, 2011 for more details). Colours are used on the plots to aid comparison.

Category	**Class**			
Vegetation (V)	None (**n**)	Separate tile (**s**)	Integrated (**i**)	
No. of models	8	19	5	
Anthropogenic heat flux (A_N)	None (**n**)	Prescribed flux (**p**)	Internal building temperature (**i**)	Modelled (**m**)
No. of models	22	2	6	2
Temporal variation of the anthropogenic heat flux (T)	None (i.e., no flux) (**n**)	Fixed (i.e., time invariant flux) (**f**)	Variable (i.e., time varying flux) (**v**)	
No. of models	22	3	7	
Urban morphology (L)	Bulk (**1**)	Single layer (**2**)	Multiple layer (**4**)	
No. of models	6	20	6	
Facets & orientation (F_O)	Bulk (**1**)	Roof, walls, road without orientation (**n**)	Roof, walls, road with orientation, no intersections (**o**)	Roof, walls, road with orientation and intersections (**i**)
No. of models	5	17	6	4
Reflections (R)	Single (**1**)	Multiple (**m**)	Infinite (**i**)	
No. of models	11	13	8	
Albedo, emissivity (A_E)	Bulk (**1**)	Two facet (**2**)	Three facet (**3**)	
No. of models	5	4	23	
Net storage heat flux (S)	Net all wave radiation (**n**)	Surface energy balance residual (**r**)	Conduction equation (**c**)	
No. of models	3	6	23	

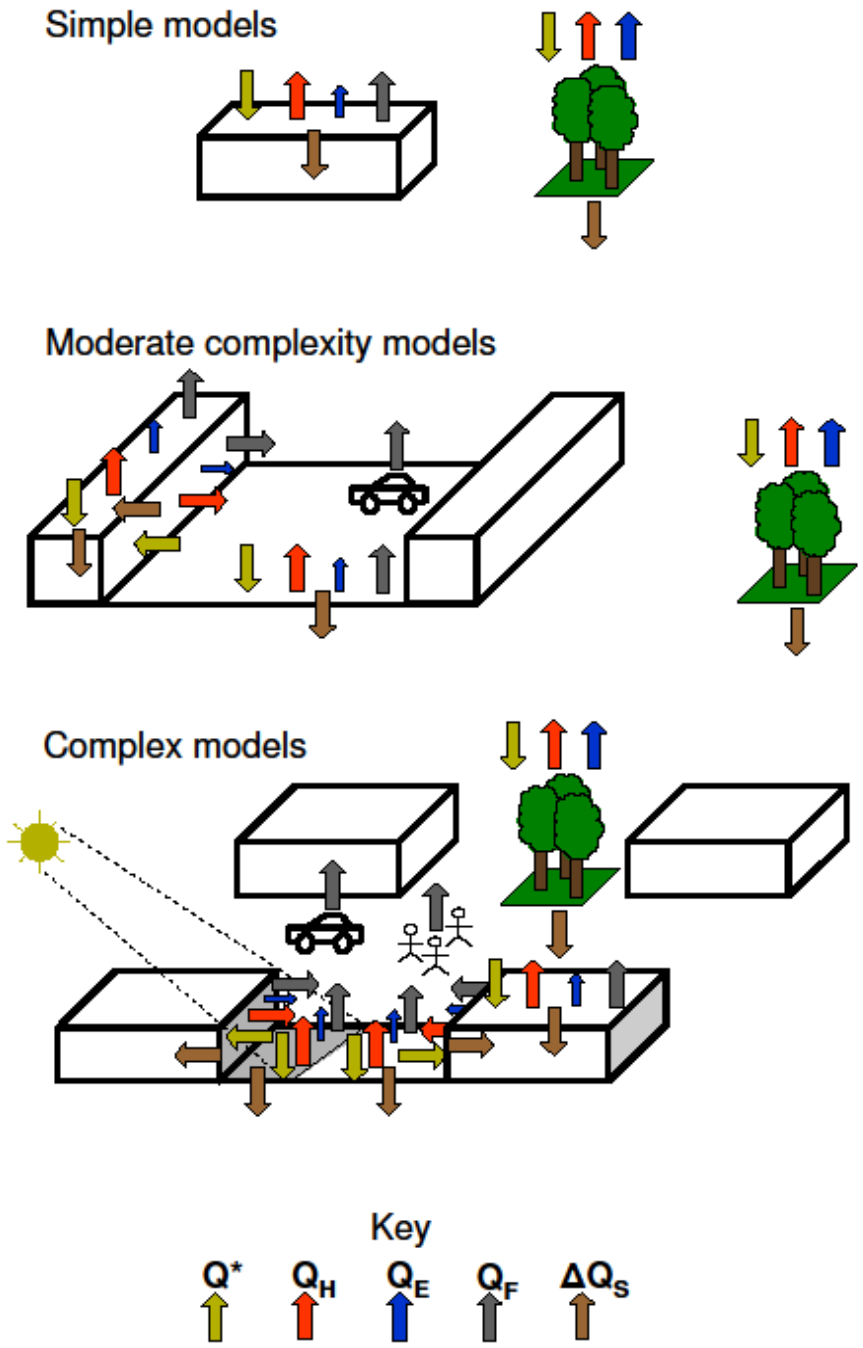

Figure 6.1: Conceptual figure of how surface energy balance exchanges are included in urban land surface models of different complexity. Note individual models have simple and complex features (Grimmond et al., 2011).

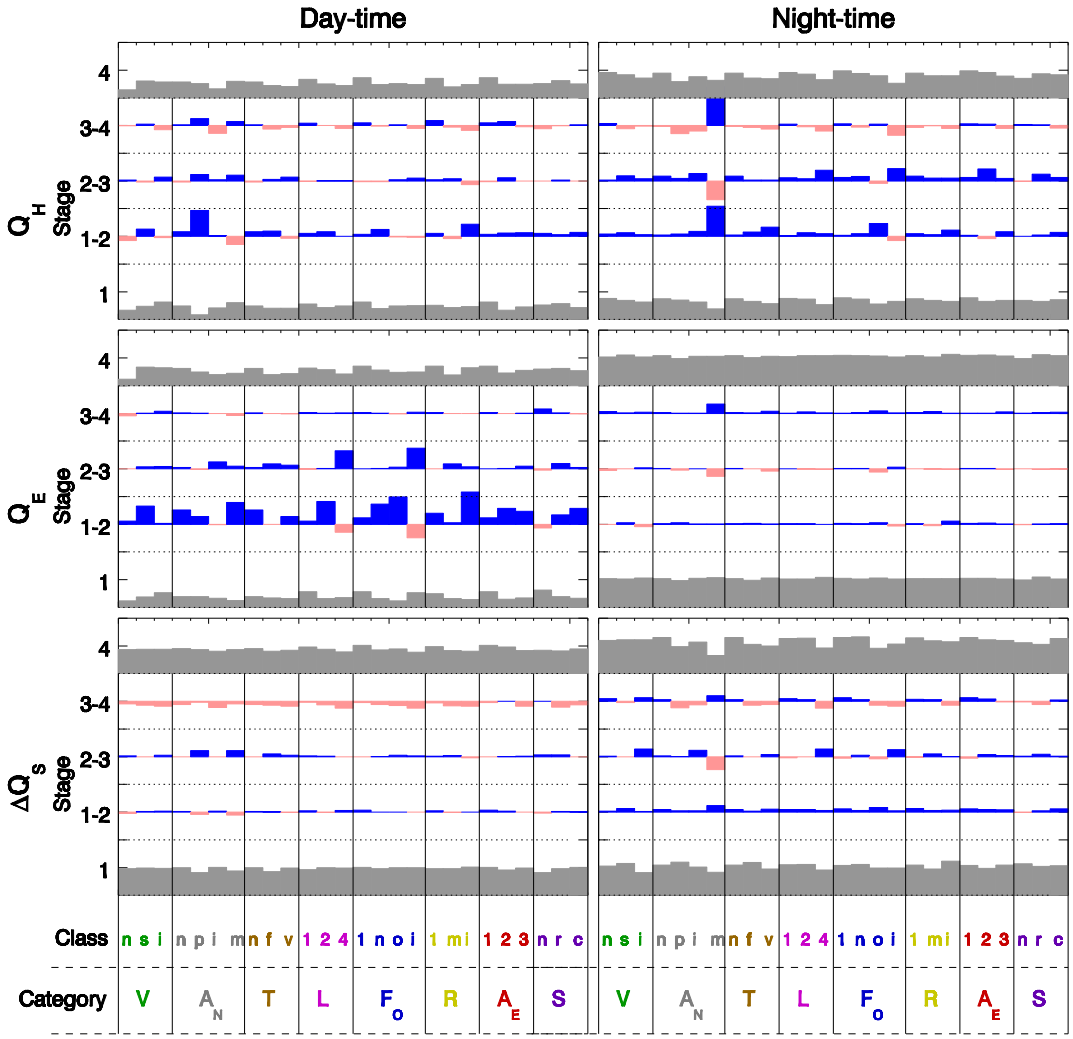

Figure 6.2: For each flux and physical category class (Table 6.4), the percentage of modelled data points within the specified observational errors (eqn. 6.1) for Stages 1 and 4 (grey) plus the change relative to the previous stage (eqn. 6.2; scaled between -100% and 100%, shown by the horizontal dotted lines). Blue shading indicates an improvement (> 0) and red degradation (< 0). Results are shown for day and night-time (with day defined as incoming solar radiation flux greater than 0 W m^{-2}). Codes definition for the physical categories and component classes (used in the x-axis) are given in Table 6.4.

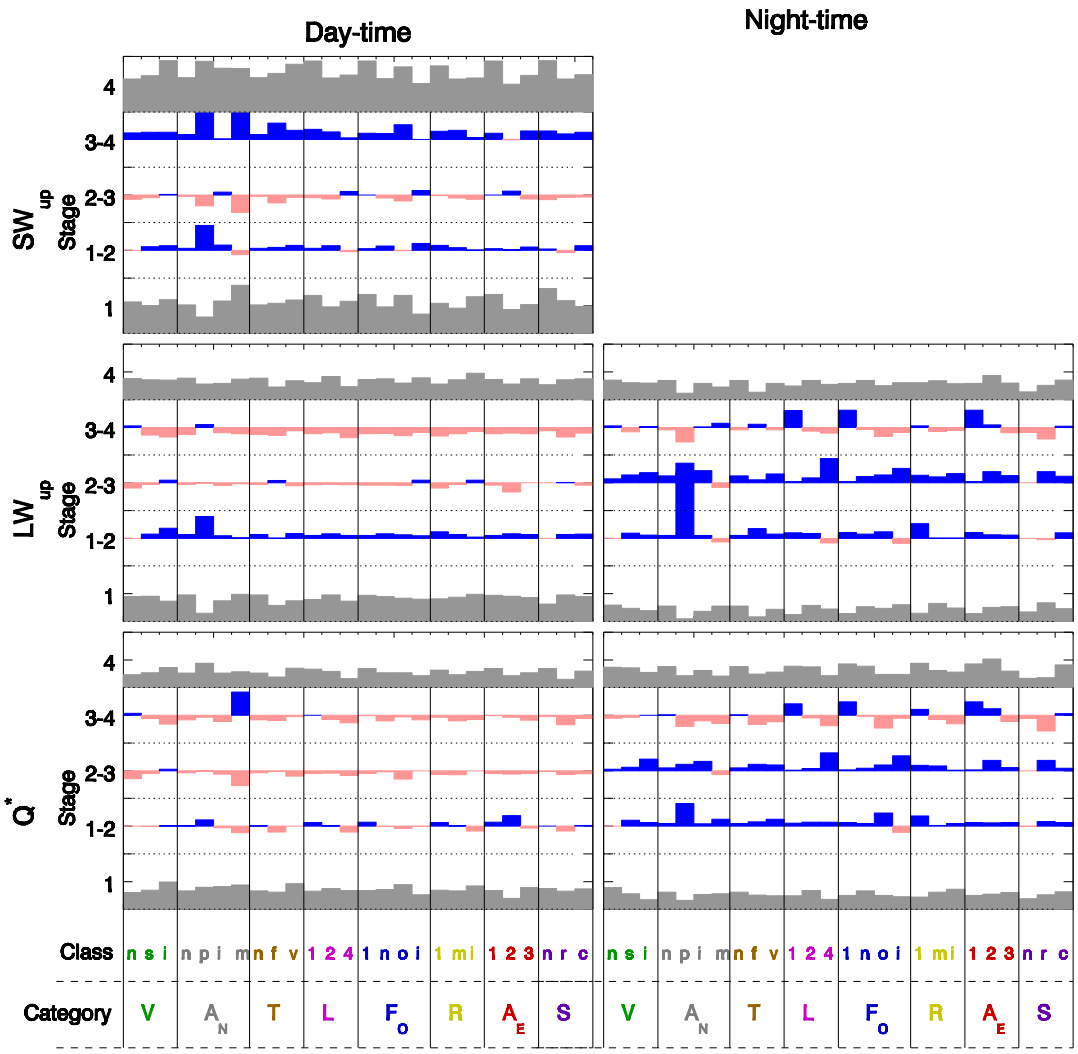

Figure 6.3: As for Fig. 6.2, but for the radiative fluxes.

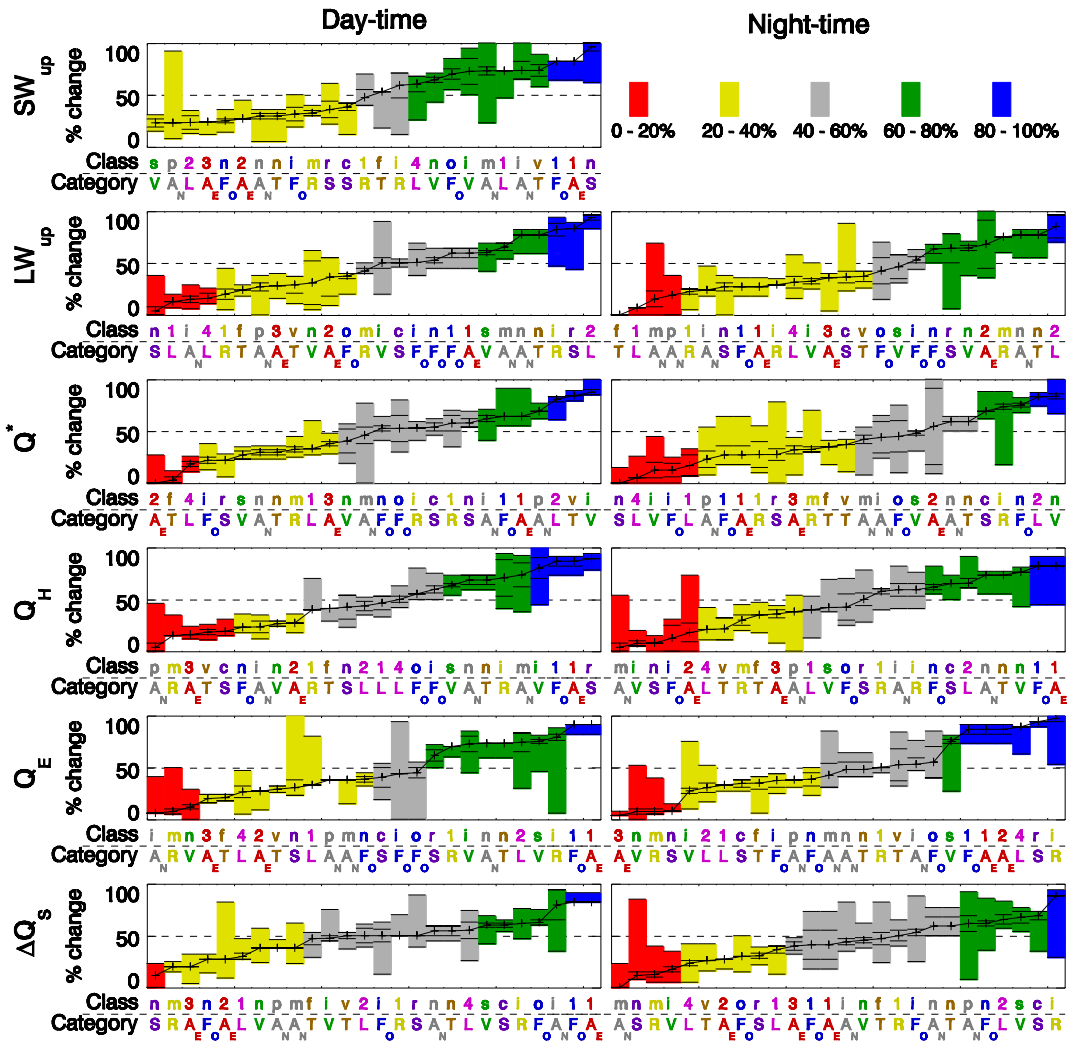

Figure 6.4: The subset of models within a class of a category improved compared to all models (P_{ca}, eqn. 6.4) ranked according to the median over the stages (for each flux, by time of day (as for Fig. 6.2)). Shading shows the range of results over the stages, with the individual results shown as horizontal lines within this. The colouring emphasises the values of the median over the stages, with 100% corresponding to all classes improved, 0% all classes degraded and 50% no change. Note X-axis code (Table 6.4) order changes between subplots because of ranking (Colour text is to aid differences to be noted).

Chapter 7: How important is the anthropogenic heat flux for the urban surface energy balance?[13]

Abstract

Results from the first international urban model comparison experiment (PILPS-Urban) suggested that models which neglected the anthropogenic heat flux within the surface energy balance performed at least as well as models that include the source term, but this could not be explained. The analyses undertaken show that the results from PILPS-Urban were masked by the signal from including vegetation, which was identified in PILPS-Urban as being important. Including the anthropogenic heat flux does give improved performance, but the benefit is small for the site studied given the relatively small magnitude of this flux relative to other terms in the surface energy balance. However, there is no further benefit from including temporal variations in the flux at this site. The importance is expected to increase at sites with a larger anthropogenic heat flux and greater temporal variations.

7.1. Introduction

Numerical weather prediction (NWP) models have included surface processes for many years (e.g., Manabe, 1969), but it is only within the last decade that representations of urban areas have been included (e.g., Best, 2005), even though urban energy balance models themselves have been developed over a much longer period (e.g., model evaluated by Ross and Oke, 1988, Masson, 2000, Martilli et al., 2002). The increasing resolution for NWP models has now reached the stage where urban areas can make up a large proportion of a grid-box, or in some instances actually be resolved. This has led to additional interest from this community to include a representation of urban areas within their modelling systems. Also, the move towards more integrated impacts for climate change has seen a move away from the post processing of urban areas from climate change signals, to including cities within the climate change simulations themselves (e.g., Oleson et al. 2008a, McCarthy et al. 2010).

Unlike natural surfaces, the energy exchange within an urban environment includes additional

[13] *This has been published as:* Best, M.J., C.S.B. Grimmond (2015), Investigation of the impact of anthropogenic heat flux within an Urban Land Surface model and PILPS-urban, *Theoretical and Applied Climatol*. doi:10.1007/s00704-015-1554-3.

source terms from the human activities (e.g., Sailor, 2011). These include the energy that is released from the heating of buildings and the emission of heat from vehicular transport. There is also a contribution from the metabolic metabolism of the humans themselves, although this is typically small compared to the other sources (Sailor, 2011). Calculations of the total magnitude of this anthropogenic heat vary considerably between cities (e.g., Christen and Vogt, 2004, Offerle et al., 2005a, Quah and Roth, 2011, Kotthaus and Grimmond, 2014a), and between different areas of any particular city (e.g., Ichinose et al., 1999, Pigeon et al., 2007b, Iamarino et al., 2012). In some locations the magnitude of the anthropogenic heat flux can be a substantial source term, similar to the daily mean solar forcing (e.g., Ichinose et al., 1999, Iamarino et al. 2012).

A priori it would be expected that such an additional source term would need to be accounted for in any urban model. However, results from the first international urban model comparison project (PILPS-Urban) consistently suggested that models which do not include an anthropogenic heat flux performed at least as well as models that did include this flux (Grimmond et al., 2011, Best and Grimmond, 2013). These studies were not able to suggest the reasons for this and indicated that additional investigation is required.

Further to the results of these studies, Figure 7.1 shows the seasonal errors derived from the results of PILPS-Urban for the median model in each group, when the models are categorised by the complexity with which they represent the anthropogenic heat flux. These have been calculated using the methodology presented in Best and Grimmond (2013). The results show that the group of models that do not include the anthropogenic heat flux have the smallest root mean square errors (RMSE) for all four of the surface fluxes at all times of the year, compared to any of the other model groups that include this additional energy flux in various forms. In addition, the models without the anthropogenic heat flux also have the smallest bias for the sensible and latent heat fluxes, although they are the only group of models that have a negative bias in the sensible heat flux during the winter months.

Here we aim to understand the counter intuitive results from PILPS-Urban. To do this we use one of the models from the comparison which did not include any anthropogenic heat flux. We have rerun the simulations that were done for PILPS-Urban whilst introducing the additional source of energy to this model and analysed the subsequent impact on the results without this term, as submitted to the comparison.

7.2. Methodology

7.2.1 Observations

The observational site chosen for PILPS-Urban was Preston, a northern suburb of Melbourne, Australia. Details of the site are given in Coutts et al. (2007b), and have also been summarised in the various analyses of PILPS-Urban results (Grimmond et al., 2011, Best and Grimmond, 2013, 2014). The site is described as urban climate zone (UCZ) 5, with moderately developed low density housing (Coutts et al., 2007b). Two methods were used to determine the surface cover fractions over a 500 m radius with the average giving 45% building area, 5% concrete, 13% roads, 15% grass, 23% other vegetation and 1% other (Coutts et al, 2007b). This gives the total impervious surface as 62% and the total pervious as 38%.

Observations of the radiative, sensible and latent heat fluxes were all taken at the site whilst the net storage heat flux was determined as the residual of the surface energy balance. The net advective heat flux is difficult to determine and was assumed to be negligible, and the anthropgenic heat flux was derived using an inventory approach. Details of the instrumentation and sampling and averaging periods are given in Coutts et al. (2007b) and are not discussed further here.

The method used to determine the anthropogenic heat flux followed Sailor and Lu (2004), utilising available data sources. This method considers the heat released from three different sources: vehicles, buildings (which is subsequently split further into electricity and natural gas), and human metabolism. To determine the contribution from vehicles, data were taken from a survey during November 2002 – October 2003 (Coutts et al., 2007b) and used along with population data from the 2001 census to determine an estimated value of the distance travelled per person per day. These data could not be used to determine hourly values, so the hourly fraction of daily traffic was taken as an average of U.S. cities from Hallenbeck et al. (1997).

For building electricity half-hourly demand data were used, but only the fraction used for direct heating was accounted for. This is 43.1% of the total electricity usage, whilst the remaining 56.9% was used for refrigeration, lighting and appliances, for which heat is only a small by-product (and is thus neglected). For natural gas the diurnal heating profile was estimated using the diurnal variability in consumption which was modelled as a function of the daily range in temperature, using mean maximum and minimum temperatures occurring at 1700 and 0700 local solar time (LST) respectively, and linear interpolation.

The human metabolic rates, along with the day-time and night-time periods, were taken from Sailor and Lu (2004), i.e., 175 W between 0700 – 2100 LST and 75 W between 2300 – 0500 LST, respectively. During the transition periods (0600 and 2200 LST) a fixed value of 125 W was used, in contrast to Sailor and Lu (2004) who used linear interpolation between the day and night-time values. These metabolic rates were then used along with the population density to determine the contribution to the anthropogenic heat flux.

The contribution of the anthropogenic heat flux from human metabolism calculated by Coutts et al. (2007b) is small compared to the other source terms, as seen in other studies (e.g., Sailor and Lu, 2004). For the other three components, the magnitude of their contribution is similar. The vehicle term has distinctive double peak during morning and afternoon rush hours, as documented by Sailor and Lu (2004), but this is to be expected because the diurnal variations were determined from the same U.S. datasets. The natural gas term peaks in the morning at the time of minimum temperature (0700 LST) and has its smallest value in the afternoon at the time of maximum temperature (1700 LST), whereas the electricity term is fairly constant throughout the day. The resultant diurnal cycle for the anthropogenic heat flux has two peaks, but with the morning peak being greater and the afternoon peak (Coutts et al., 2007b).

7.2.2 Urban land surface model

The Joint UK Land Environment Simulator (JULES, Best et al., 2011), used in the current study, provided four sets of results to PILPS-Urban. There were two configurations, which represented the urban fraction as a single bulk surface, or the roof surface and street canyons separately; namely the 1-tile and 2-tile versions (see Best 2005 and Best et al. 2006). These two configurations were run by two modelling centres and ensured that the physical set up of the models was consistent, but the assumptions about the initial conditions, especially the soil moisture, were different. One set of simulations had more initial soil moisture than the other, which was shown by Best and Grimmond (2014) to have important implications for the model performance. None of the four sets of JULES simulations included an anthropogenic heat flux.

Despite the differences in physical configuration and initial conditions between the four sets of JULES simulations, results from PILPS-Urban show that all of them perform well compared to other models (Figure 7.2, adapted from Grimmond et at., 2011), especially for the sensible and latent heat fluxes. Hence this is a good model to use to investigate the impact of including the additional anthropogenic heat flux.

For the simulations presented here, both the 1-tile and 2-tile versions of JULES were used, but with the initial soil moisture set to the same values as used for the dryer set of results contributed to PILPS-Urban.

To put the results for the impact of including the anthropogenic heat flux into context with other aspects of the physical system represented in the urban models, an additional simulation using JULES with no representation of vegetation was undertaken. Results from the previous studies of PILPS-Urban concluded that a representation of vegetation was critical in order to obtain good performance from the urban models, especially for the sensible and latent heat fluxes (Grimmond et al., 2010, 2011, Best and Grimmond, 2013, 2014). However, the method by which the vegetation is represented, i.e., though an independent surface (tile scheme) or interacting with the urban surface (integrated) was shown to be less important.

The JULES model included a tile scheme representation for vegetation in all four of the simulations submitted to PILPS-Urban, and the same representation has been used in all of the simulations with the anthropogenic heat flux. However, an additional simulation was completed with no anthropogenic heat flux and with the vegetation removed. This was done by setting the fraction of the vegetation surface to zero and re-scaling the urban surface fractions to sum to unity.

7.2.3 Anthropogenic heat flux

The JULES model was adapted to include the anthropogenic heat flux as an additional term in its surface energy balance:

$$C\frac{\partial T}{\partial t} = Q^* - Q_H - Q_E - \Delta Q_S + Q_F \qquad (7.1)$$

where C is the areal heat capacity of the surface, T is the surface temperature, Q^* is the net all-wave radiation, Q_H is the turbulent sensible heat flux, Q_E is the latent heat flux, ΔQ_S is the net storage heat flux and Q_F is the anthropogenic heat flux.

The anthropogenic heat flux was specified at every time-step of the model run, based upon the observed values of the anthropogenic heat flux. By using the observed values, the analysis removes the impact of using a scheme to represent the flux which would

inevitably have its own inaccuracies. Hence we can identify the true impact of including the anthropogenic heat flux.

For the 1-tile version, the anthropogenic heat flux was applied to just the urban surface energy balance. However, for the 2-tile version there is more flexibility regarding the addition of the flux. It can be added to just the canyon surface energy balance, the roof surface energy balance or to both the canyon and the roof surface energy balances. For the purpose of understanding the full impact of the anthropogenic heat flux on the JULES simulations, the 2-tile version has been run using all three of these combinations. For each of these, the anthropogenic heat flux applied to each surface has been scaled to ensure that the total flux integrated over all surfaces is equal to the observed values.

In addition to investigating the impact of including the anthropogenic heat flux, we also study the results of including temporal variation in the flux. Four methodologies to provide the time-step values of the flux were used (Fig. 7.3a-d), namely:

1. The average value of the period of the observations: constant (i.e., no temporal variation)
2. Monthly mean values, constant diurnal cycle, causing a step change between consecutive months
3. Average diurnal cycle over the entire observational period, constant variation between months
4. Monthly mean diurnal cycle, with variations between months (estimated Q_F data, Coutts et al, 2007b)

The latter is the full temporal resolution of Q_F. Note that there are no differences between the diurnal cycle of days within the same month, due to the methodology used to determine the observed values; i.e., no response to meteorological conditions or human behavior (e.g., days of the week).

The average value of the anthropogenic heat flux in the observational dataset is 11 W m^{-2}, but as this is only applied to the urban surfaces of the land cover in JULES (i.e., not applied to vegetation or bare soil surfaces) the average value of the anthropogenic heat flux used in JULES is 17 W m^{-2}. This gives a diurnal maximum value of 26 W m^{-2} and a minimum of 10 W m^{-2}. Figure 7.3e shows the average diurnal cycle over the observational period, along with the diurnal cycles for the months with the maximum and minimum values. This figure shows two important aspects for understanding the results presented here. Firstly, the magnitude of the flux

is small compared to the average of the other terms in the surface energy balance (Q^* = 83 W m^{-2}, Q_H = 83 W m^{-2}, Q_E = 34 W m^{-2} and ΔQ_S = 20 W m^{-2}), and in particular to the average downward component of shortwave radiation ($K\downarrow$ = 168 W m^{-2}). Secondly, the minimum values in the diurnal cycle occur during the night-time hours, when the sensible heat flux is small.

7.2.4 Analysis methods

The mean bias error (MBE) and root mean square error (RMSE) for Q^*, Q_H and Q_E for each of the simulations are presented in Figure 7.4, for all data points and separately for the night-time values only (defined by $K\downarrow$ = 0.0 W m^{-2}). The statistics for the net storage heat flux are not shown as this is taken to be the residual of the energy balance in the observations and as such aggregates the observational errors. To calculate these statistics, any time-steps with missing observational data for any flux are omitted from the analysis. This is to ensure that the results are consistent between the fluxes, and with the methodology that was adopted in PILPS-Urban by Grimmond et al. (2011) and Best and Grimmond (2013, 2014).

7.3. Results

Figure 7.4 shows that the JULES simulations that include the anthropogenic heat flux have better performance (smaller MBE and RMSE) for the sensible heat flux, compared to the simulations without the anthropogenic heat flux (as submitted to PILPS-Urban). This holds for the analysis using all of the data and using the night-time data only. The latent heat flux has a very small increase in the positive bias over all of the data and a decrease in the negative bias for the night-time only data, but virtually no change to the RMSE. However, for Q^* there is an increase in the MBE for both all data and the night-time data and a corresponding increase in the RMSE.

The increase in negative bias for the night-time results for Q^* are a result from the higher surface temperatures leading to more outgoing longwave radiation. However, since Q_H still has a negative bias when Q_F is included, this suggests that the relation between the radiative surface temperature and the thermodynamic temperature within JULES is not optimal and could be improved.

The results show that for the anthropogenic heat flux with temporal variations there is little impact on all three of the fluxes (Figure 7.4), with only small differences in either the MBE or the RMSE. Hence for these simulations, including the average value of the anthropogenic heat

flux is more important than having time varying values, either diurnally of monthly. However, it should be noted that for this study only the mean monthly variations in Q_F are available and not the true temporal variability that depends on the meteorological conditions and human behavior, such as weekdays versus weekends.

All of the 2-tile versions perform better than the equivalent 1-tile version for Q* and Q_H for both MBE and RMSE for all data and night-time data only. The only exception is for the MBE in Q_H with the simulations that do not include vegetation. There is no difference between the 1-tile and 2-tile versions for Q_E for either the MBE or the RMSE, except for the night-time MBE for the 2-tile version with the anthropogenic heat flux only applied to the canyon surface, which has a slightly larger negative bias.

There are also differences with the 2-tile version of JULES for Q* and Q_H when considering the surfaces over which the anthropogenic heat flux are implemented. The simulations with the anthropogenic heat flux applied to just the canyon surface have noticeably smaller MBE than the other simulations for both all data and night-time only data, but these improvements are not evident in the RMSE apart from the night-time data for Q_H which has a very small improvement. Consistent with these results, we find that not applying the anthropogenic heat flux to the canyon tile (i.e., applying it only to the roof tile) leads to larger MBE and RMSE, particularly for the sensible heat flux.

Despite the improvements that can be detected from the model simulations that include the anthropogenic heat flux, these are far smaller than the improvements that are obtained from including a representation of vegetation, especially for the RMSE for all of the data. These results are robust over all of the sets of simulations.

The maximum values of the anthropogenic heat flux occur in the winter months (Fig. 7.3a), when the diurnally averaged sensible heat flux is at its lowest values. Hence we might expect to see a larger impact from the anthropogenic heat flux on Q_H at this time of the year. Seasonal variations of 60 day means in the results for the 2-tile version of JULES, with the anthropogenic heat flux applied only to the canyon energy balance, are shown in Figure 7.5 for the night-time. This figure shows the mean flux, MBE and the RMSE for all of the surface fluxes and is equivalent to the analysis presented in Best and Grimmond (2013) for all of the models in PILPS-Urban. For both the day-time (not shown) and night-time results (Fig. 7.5), the seasonal cycle of the anthropogenic heat flux generally makes no difference to the improvement of the model in terms of MBE for any of the fluxes. The results are consistent across all of the months,

with the only exception being a slight improvement to the MBE of ~1.5 W m^{-2} for the night-time Q_H during June/July (JJ) compared to the summer months. However, there is a slightly larger improvement in the RMSE for Q_H (~ 3 W m^{-2}) in both the day-time (not shown) and night-time (Fig. 7.5) during the winter months (JJ) compared to not including the anthropogenic heat flux. For nocturnal data this improvement results from the RMSE not increasing by as much as for the results from the model without anthropogenic heat flux.

For the net storage heat flux, there is a slight improvement in both the MBE and RMSE during the day-time from including the anthropogenic heat flux (not shown), but a degradation in both statistics for the night-time (Fig. 7.5). These changes to both MBE and RMSE are consistent throughout the seasonal cycle.

As well as the anthropogenic heat flux having larger impacts at certain times within the seasonal cycle, we might expect the impact on the surface fluxes to vary during the diurnal cycle, especially when Q_H has its smallest values (i.e., during the night-time). However, an equivalent figure to Fig. 1 of Best and Grimmond (2013), for the average diurnal cycle for each 60 day period of the seasonal cycle, shows that the impact on the diurnal cycle for all of the surface fluxes is small (not shown). At the scale of the range of the diurnal cycle, the differences in the various model simulations that include the anthropogenic heat flux are not discernible from those without an anthropogenic heat flux.

Focusing on just the night-time part of the diurnal cycle (Fig. 7.6), results show that for this period there are some noticeable differences in the sensible and net storage heat fluxes between the model simulations. The JULES runs that include the anthropogenic heat flux show a reduced negative bias during the night-time compared to the run without this flux. This reduction in the negative bias ranges from a minimum of 4.8 W m^{-2} in the summer to a maximum of 7.3 W m^{-2} in the winter.

For the net storage heat flux, there is a reduction in the magnitude of the negative flux during the night-time from the runs with Q_F compared to the run without Q_F. The results for the JULES simulation without an anthropogenic heat flux have only a small bias in the net storage heat flux at night-time over most of the seasonal cycle. So subsequent changes to the night-time values in the net storage heat flux from implementing an anthropogenic heat flux into JULES introduces a more notable positive bias at all times of the year.

7.4. Conclusions

The JULES model was one of the better performing models in PILPS-Urban, especially for Q_H and Q_E. However, this model did not include any representation of the anthropogenic heat flux within the simulations for the comparison. Hence it is a good tool for investigating if the inclusion of Q_F could improve the model performance still further.

The results from the runs presented here to investigate this have shown that there is an improvement in both the MBE and RMSE throughout the seasonal cycle from including the anthropogenic heat flux in JULES. Whilst the improvement to the MBE is fairly constant throughout the year for the day-time results, the improvement in the night-time MBE and the day-time MBE and RMSE is greater in the winter months, when Q_H has its smallest average diurnal values, compared to the summer months.

A positive impact is also evident for Q_H during the night-time, with the consistent negative bias from the simulations without Q_F reduced when the Q_F term is included in the surface energy balance of JULES. These changes also lead to a greater reduction in the night-time RMSE in the winter months than in the summer, as might be expected since Q_H has its lowest values.

The impact of including Q_F on Q* is not so beneficial, leading to a slight degradation in both the MBE and RMSE. However, this negative impact of Q_F within the JULES simulations is more indicative of issues with the radiation balance rather than a direct influence of Q_F itself. It is possible that if the radiation issues within JULES were improved, the impact of including Q_F might also give beneficial results for Q*.

There are also negative impacts on the errors for the net storage heat flux, with both the MBE and RMSE being degraded when Q_F is included in the simulations. The negative bias in ΔQ_S during the day-time throughout the year, along with the positive night-time bias, suggests that insufficient energy is being stored during the day-time and subsequently released during the night-time. This is consistent with the results presented by Best and Grimmond (2014) who suggested that urban models have energy partitioning issues in general between Q_H and ΔQ_S.

Whilst the inclusion of the anthropogenic heat flux leads to some improvements within the simulations, the magnitude of these improvements is small, even to the extent that it is difficult to identify the changes when looking at the full diurnal cycle of the fluxes. The magnitude of the anthropogenic heat flux within the observational dataset has an average value of 11 W m^{-2}, which is typical of suburban areas. This could suggest that Q_F is not significant in these

environments, especially since the minimum values of Q_F occur at similar times during the night to the smallest values of the sensible heat flux. However, the temporal variations in Q_F used in this study were only the mean monthly variations and do not take into account the meteorological conditions, or human behavior such as weekday and weekend activities, hence the actual variations in Q_F might have a greater impact. In addition, for urban areas in colder climates, Q_F could contribute a relatively larger fraction to the surface energy balance, due to the smaller size of radiative fluxes, and hence be more significant. The flux is also known to be much greater in dense, urban centres (e.g., Ichinose et al., 1999).

The seasonal variations in Q_F are also small, which explains why the increased positive impact from including the flux during the winter months is also small. The small magnitude and diurnal cycle of the anthropogenic heat flux at this site could well be responsible for this result. It is likely that at sites with large variations in both the diurnal and seasonal cycles in Q_F there will be additional benefits from resolving the temporal behavior of the anthropogenic heat flux.

The impact of including Q_F is much smaller on both the MBE and RMSE for all of the data and night-time data only, than including a representation of vegetation for the site. Again this result is influenced by the relatively small magnitude of the anthropogenic heat flux. However, nearly all of the models that neglected Q_F in PILPS-Urban did include a representation of vegetation, whilst other categories that included Q_F contained a greater proportion of models that neglected vegetation. Hence the counter intuitive results presented in the PILPS-Urban, suggesting that the group of models that did not include the anthropogenic heat flux performed at least as well as the models that did include this flux, were being influenced by the treatment of vegetation within these models.

From the results presented in this study we can conclude that a representation of the anthropogenic heat flux is important for urban models and can lead to improved results. Moreover, the influence of the anthropogenic heat flux is likely to be greater at sites with a larger flux, increasing the need for the urban models to include this term in their surface energy balance.

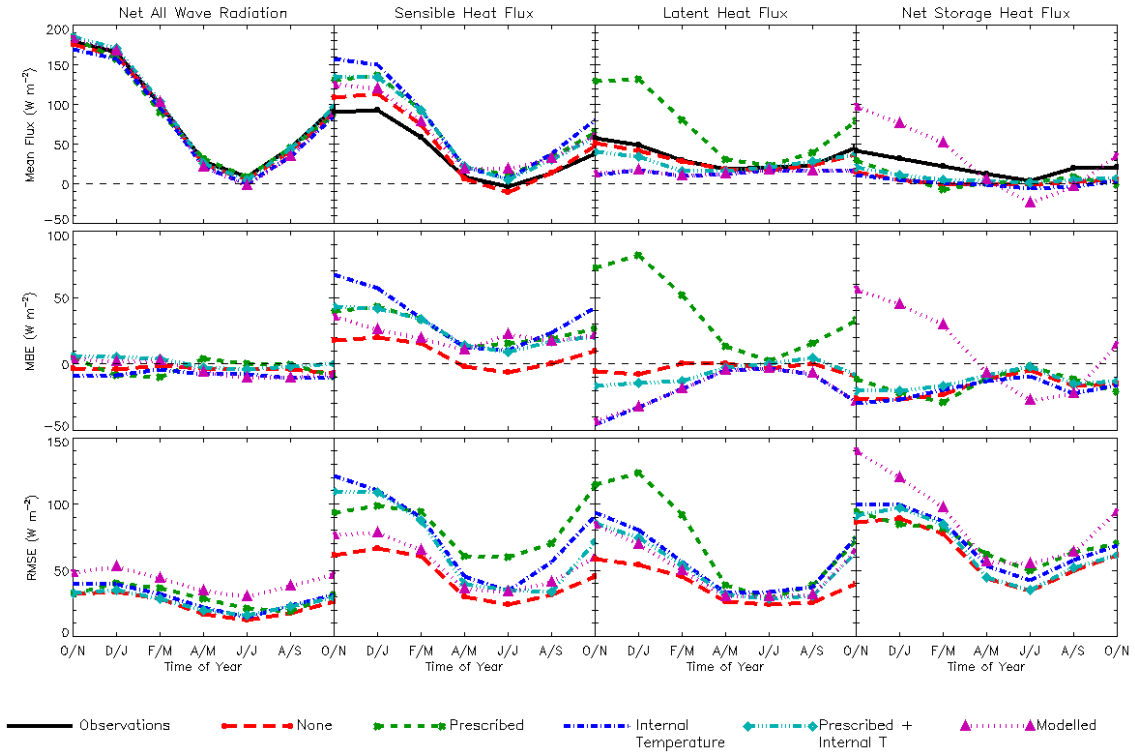

Figure 7.1: Median of the mean modelled flux, mean bias error (MBE), and root mean square error (RMSE) for the surface fluxes from the models in PILPS-Urban, determined for two month periods, for the models classified by their representation of the anthropogenic heat flux. Note the scales are different for each flux. For details of PILPS-Urban see section 2, and for method of analysis see Best and Grimmond (2013).

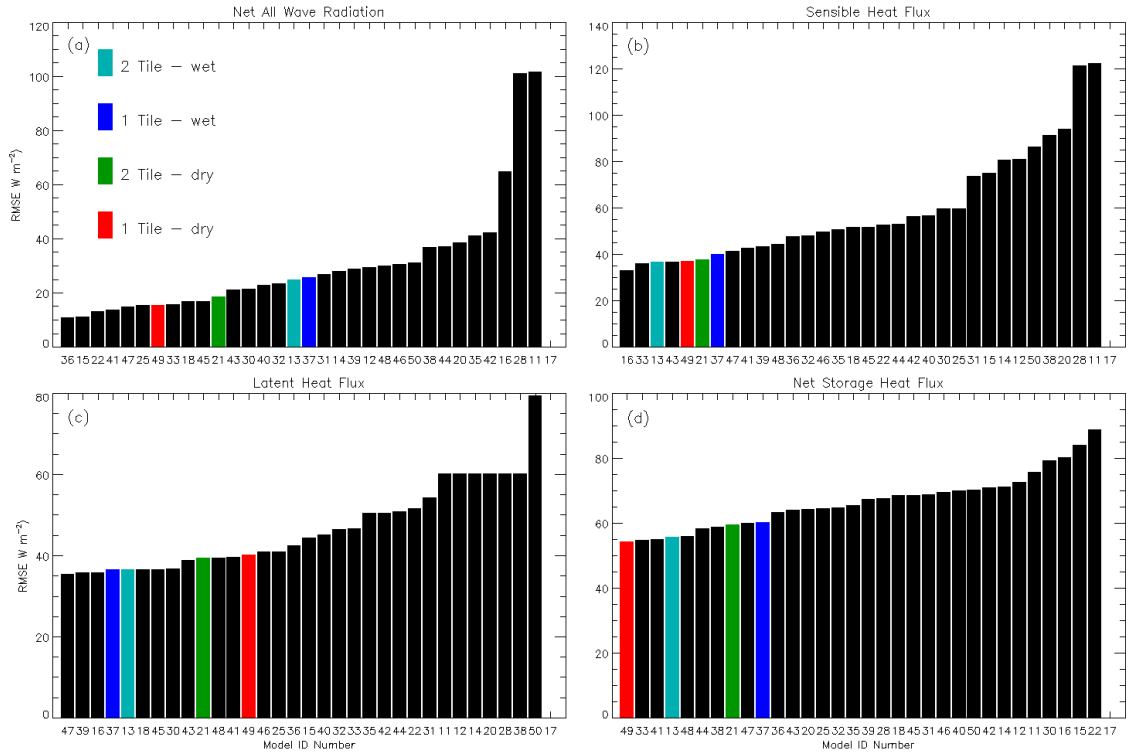

Figure 7.2: Ranked model performance for stage 4 of PILPS-Urban, for each of the surface fluxes. The four sets of results from the JULES model are highlighted. These were generated from the 1- and 2-tile configurations, each run with low (dry) and high (wet) initial soil moisture.

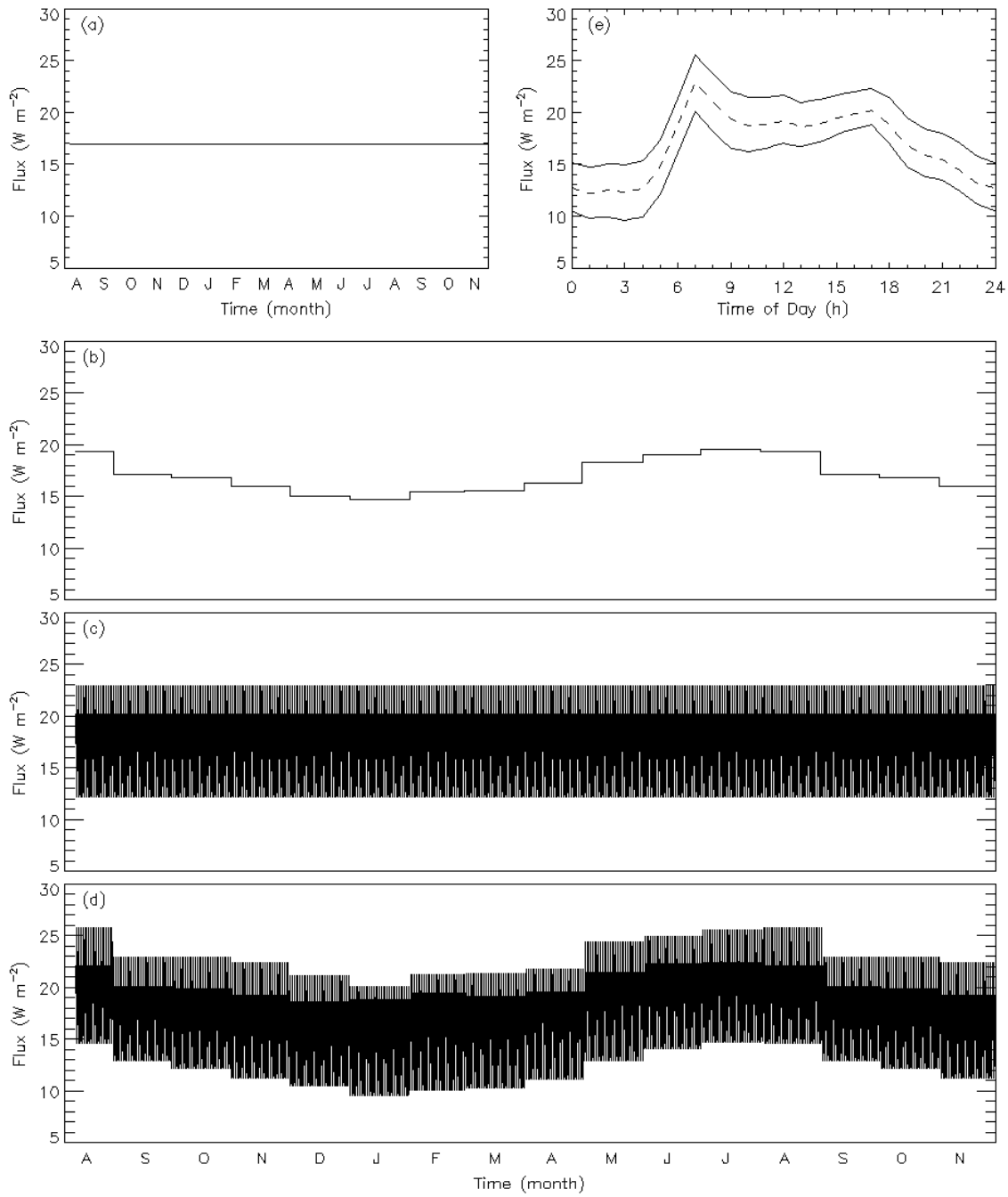

Figure 7.3: Temporal variations in the anthropogenic heat flux applied to the urban surface fraction for the model simulations: (a) mean for observational period, (b) mean monthly value, (c) mean diurnal cycle for observational period, (d) monthly mean diurnal cycle, (e) mean diurnal cycle for observational period applied to the urban surface fraction (dashed line), with the average diurnal cycles for the months with maximum and minimum values (solid lines).

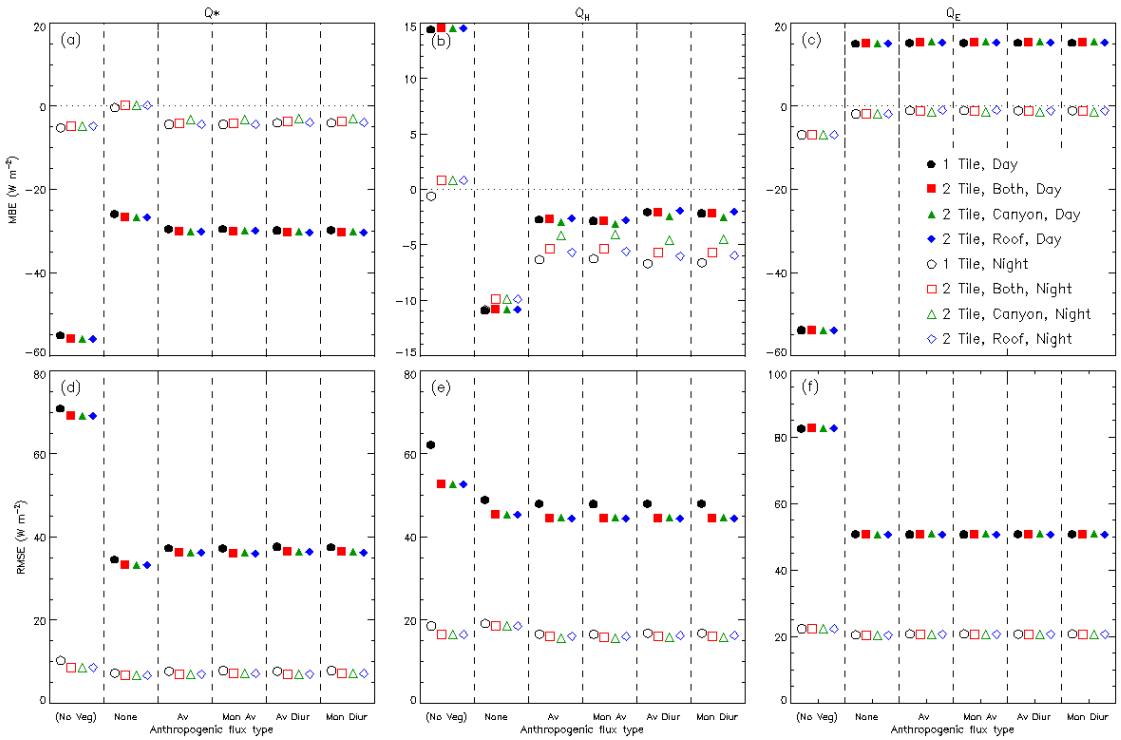

Figure 7.4: MBE (a-c) and RMSE (d-f) for net all-wave radiation (a,d), sensible heat flux (b,e) and latent heat flux (c,f), for both day- (filled) and night-time (hollow), for all implementations of temporal variation in the anthropogenic heat flux, and simulations without vegetation. Results are presented for the 1 and 2 tile configurations with the anthropogenic heat flux applied to only the canyon, only the roof, or both canyon and roof. Note the scales are different for each flux.

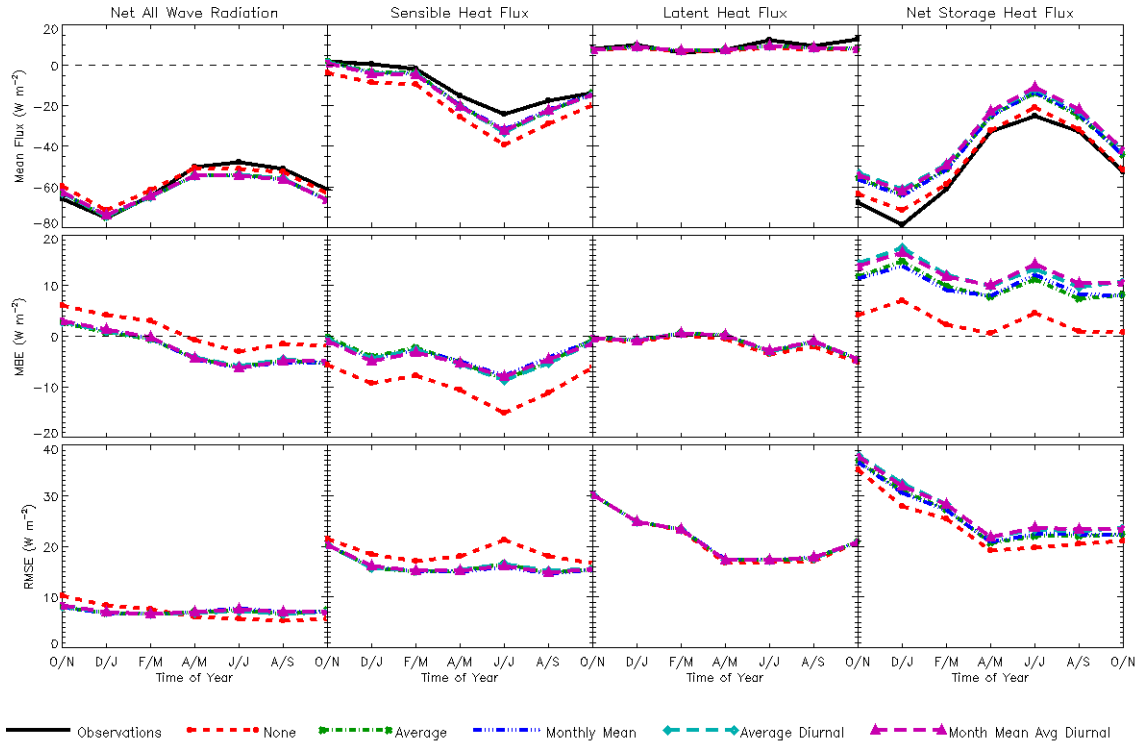

Figure 7.5: Nocturnal mean modelled flux (row 1), MBE (row 2), and RMSE (row 3) for the surface fluxes determined for two month periods.

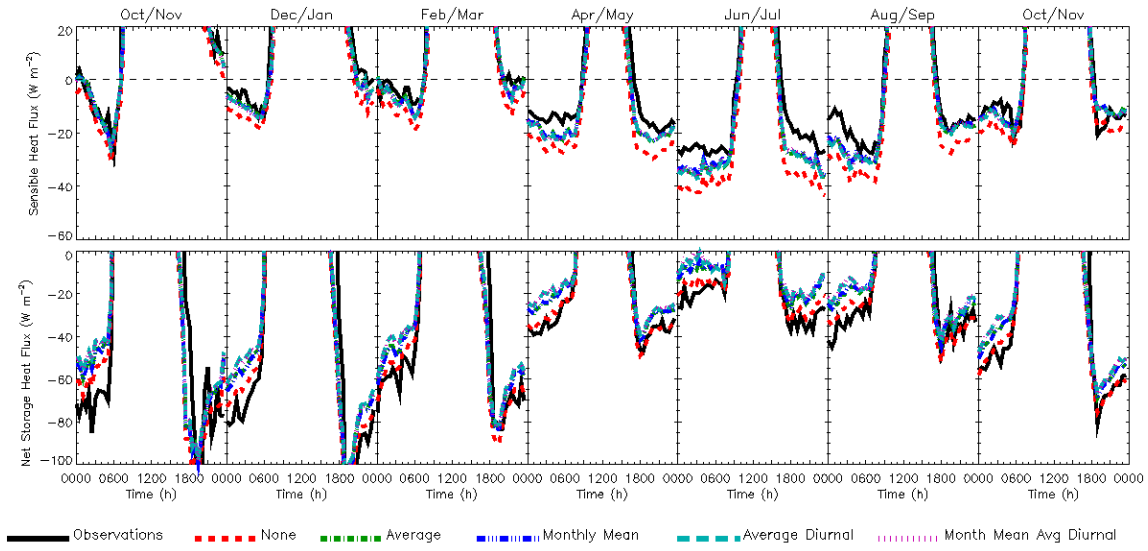

Figure 7.6: Mean diurnal cycle for each 60-day period throughout the seasonal cycle, scaled to focus on the night-time results, for the sensible and net storage heat fluxes. Note the scales are different for both fluxes.

Chapter 8: Modelling the Bowen ratio at a number of urban sites over a range of vegetation cover[14]

Abstract

Inclusion of vegetation has been found to be critical for urban land surface models (ULSM) to be able to reasonably represent the turbulent sensible and latent heat flux densities in an urban environment. Here the Joint UK Land Environment Simulator (JULES), a ULSM, is used to simulate the Bowen ratio at a number of urban and rural sites with vegetation cover varying between 1% and 98%. The results show that JULES is able to represent the observed Bowen ratios, but only when the additional anthropogenic water supplies into the urban ecosystem are considered. For example, the impact of the external water use (irrigation, street cleaning) on the partitioning of energy within the surface energy balance can be as substantial as that of the anthropogenic heat fluxes on the sensible and latent heat fluxes. The Bowen ratio varies from 1-2 when the plan area vegetation is in between 30% and 70%. However, when the vegetation cover is below 20% the Bowen ratios tend to increase substantially (2-10) and have higher degree of sensitivity to assumptions about external water use. As this type of area (vegetation cover < 30%) still have few log term measurements, there is a clear need for more long term observational studies.

8.1. Introduction

Over the last couple of decades there have been a number of models developed to represent urban land-surface-atmosphere interactions, such as Martilli et al. (2002), Fortuniak (2003), Kondo et al. (2005), Oleson et al. (2008a), and Salamanca et al. (2010). Typically these models are designed to represent the energy balance of the various facets that make up an idealized urban canopy. Often this is a symmetric street canyon geometry with varying degrees of complexity, ranging from a bulk canyon (e.g., Best, 2005), separate roof, walls and road, with single (e.g., Masson, 2000) or multiple (e.g., Krayenhoff and Voogt, 2007) energy balances and even intersections separately from street canyons (e.g., Kawai et al., 2009). Whilst this may be a good representation of the central downtown areas of major cities, this design alone does not capture the influence of vegetation that is typically present in large amounts at the more

[14] *This has been submitted as:* Best, M.J., C.S.B. Grimmond, A. Christen (2015), Modelling the partitioning of turbulent fluxes at urban sites with varying vegetation cover, *submitted to J. Hydro-meteorol.*

suburban locations. Even downtown areas can include non negligible amounts of vegetation within the street canyons.

The implications are that vegetation also needs to be modelled for urban areas. Indeed, the first international urban model comparison experiment (PILPS-Urban) concluded that for both sites considered, the inclusion of vegetation was critical for a good simulation of the sensible (Q_H) and latent heat densities (Q_E) (Grimmond et al., 2010, 2011, Best and Grimmond, 2013, 2015). Models that included a representation of vegetation performed much better in simulating Q_H and Q_E than models that neglected it, although PILPS-Urban also concluded that the way in which the vegetation was modelled, i.e., as a separate independent surface (e.g., Dupont and Mestayer, 2006) or integrated within the urban street canyon (e.g., Lee and Park, 2008), was not as important. However, the main focus of PILPS-Urban results was a suburban site, so it is not clear how robust these conclusions are for other sites with varying percentages of vegetation within the footprint of the observations.

Observational data have quantified directly Q_H and Q_E and hence how the Bowen Ratio β (i.e. $\beta = Q_H/Q_E$) varies with the vegetation fraction across a range of values (Grimmond and Oke, 2002, Loridan and Grimmond, 2012a). Here we investigate if an urban model that includes a representation of vegetation can reproduce the observed behaviour.

In this study we use the JULES model (Best et al., 2011) that has been shown to perform well compared to other models within PILPS-Urban for simulating Q_H and Q_E (Best and Grimmond, 2015). The model is used to simulate β for urban areas that range in plan area vegetation fraction cover from 1% to 98%, that correspond to 22 observational dataset footprint descriptions.

8.2. Methods

8.2.1 Observational sites

For this modelling study, data from 22 observational tower sites (Table 8.1) are used where Q_H and Q_E were measured continuously in the inertial sublayer above usually uniform urban canopies by means of eddy covariance. The measurements represent the neighborhood-scale surface energy balance. The datasets are both short (< two months) and longer (> 12 months) duration. The longest (BA01_06) is six years. Most of the shorter duration datasets were collected in the summer months. For mid-latitudes and semi-arid climates, the summer months

are the periods during which the vegetation is most likely to experience soil moisture stress and hence limited transpiration. The datasets with observations collected during the winter (OU03 and ME93) are tropical or sub-tropical climates where the precipitation is typically less during the winter months. Hence these are also the periods that are more likely to have soil moisture stress on the vegetation. A couple of rural sites outside of Basel (B02R1 and B02R2) were added to complete the spectrum of vegetation fractions modelled.

The surface characteristics affecting the measurements have a range in vegetation cover fractions over the datasets (Table 8.2), varying from almost total vegetation cover to only 1%. For most of the sites the plan area proportions of vegetation and impervious surfaces (streets and buildings) combine to account for around 95% of the total area, with the exceptions of Tucson (TU90) and Ouagadougou (OU03). These two sites have a substantial fraction of bare soil or unmanaged land (which is modelled as bare soil), being 17% and 30% respectively. However, these sites are both in relatively dry climates and so bare soil evaporation is unlikely to have a substantial contribution to Q_E, and hence β. As such, presenting results compared to the vegetation fraction is more appropriate than comparing them to the impervious fraction.

Further details on the measurement sites can be found in the references provided in Table 8.1 for each of the sites.

8.2.2 Model description

The model used for this study was the community land surface model JULES (Joint UK Land Environment Simulator, Best et al., 2011). This model uses a tiled approach to represent different surface types and by default includes five types of vegetation (broadleaf trees, needle-leaf trees, C3 grasses, C4 grasses and shrubs) and four non-vegetation types (urban, lakes, bare soil and permanent land ice), for which the urban tile represents the impervious surfaces of an urban environment. For the vegetation surfaces the leaf area index (LAI) can vary temporally, but for this study they have been held fixed at their default values (Best et al., 2011). Soil processes are modelled using four discrete layers that have increasing thickness with depth, the layer depths being 0.1 m, 0.35 m, 1.0 m and 2.0 m respectively. Results from the PLUMBER (PALS Land sUrface Model Benchmarking Evaluation pRoject) community experiment found the performance of this model for Q_H and Q_E, at a number of sites with natural surfaces, is comparable to many other land surface models (LSM) (Best et al., 2015).

Within this model there are three ways in which the impervious urban surface (i.e., ground (e.g.,

roads, parking lots) and buildings, excluding the vegetation and bare soil) can be represented, namely the one tile (Best, 2005), two tile (Best et al., 2006) and MORUSES (Porson et al., 2010) schemes. The one tile scheme represents an urban area as a bulk surface with effective parameters. The two tile scheme separates out the roofs of buildings from the street canyon, also using effective parameters for each of the surfaces. Finally the MORUSES scheme is similar to the two tile scheme except that the parameters and canyon turbulence is parametrised and depends upon the morphology of the urban areas.

Results for all three versions were included in PILPS-urban, although MORUSES was an early version that did not include vegetation, whilst results from the one tile and two tile versions of JULES were submitted by two modelling groups with different assumptions about their initial conditions, and in particular their initial soil moisture for the vegetation. Results presented in Best and Grimmond (2014) showed that the initial conditions for soil moisture can have a substantial impact on Q_H and Q_E. However, Best and Grimmond (2015) show that all four versions of JULES performed well in simulating Q_H and Q_E compared to other models in PILPS-Urban. Hence for this study we have chosen to use solely the two tile urban scheme within JULES, but have ensured that the initial conditions for the runs are appropriate by undertaking a spin-up simulation, as described below. The performance of this model in studies of both natural surfaces and urban environments indicates that this model is appropriate for this study.

8.2.3 Spin-up strategy

For each observational site (Table 8.1) JULES was run for at least 10 years prior to the initial observations. Following the PLUMBER method (Best et al., 2015), the initial soil moisture state was specified as being saturated conditions, then the LSM was run for 10 years to ensure the soil moisture reaches a correctly spun-up state.

Given the many short duration datasets, the 0.5^{o} resolution global WFDEI (WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Weedon et al., 2014) was used to permit a 10 year spin-up. The 1979 to 2012 dataset uses ERA-Interim re-analyses (Dee et al., 2011) to downscale monthly observations from the Climate Research Unit (CRU, New et al. 1999, 2000, Harris et al., 2013) to a temporal resolution of three hours (see Weedon et al., 2011, 2014, for more details). For precipitation, WFDEI has an alternative that is derived from the Global Precipitation Climatology Centre (GPCC, Schneider et al., 2013) for the monthly observations rather than CRU. In this study the WFDEI dataset based on GPCC precipitation are used.

For each site the WFDEI grid data were extracted and used to force JULES with a 30 min time-step. The interpolation from the three hour used a simplified Sheng and Zwiers (1998) algorithm within JULES. For radiation and precipitation data, a backward time averaging (i.e., time averaging that is valid at the end of the time period) that conserves the mean quantity is used, whilst for the other forcing variables a linear interpolation is used. The WFDEI temperature and humidity data are provided at screen level whilst the wind data are at a height of 10 m. However, the JULES model has an effective surface rather than the actual surface, i.e., the displacement height is not explicitly represented, so the WFDEI data have been used to force the model without any changes to the height. This is acceptable because the spin-up only needs to be in agreement with the previous mean climate, which can still be obtained from forcing at this height.

8.2.4 Gap filling for forcing data

To drive JULES during the analysis period, processed observational datasets from Loridan and Grimmond (2012b) were used. Any gaps in these data, or for additional sites, were filled using the WFDEI data.

Given the global nature of the WFDEI data (Weedon et al., 2011, 2014) it is quite likely that there are inconsistencies for individual sites. In particular, the long term grid mean may not match a specific observational site mean. To asses this, periods with observed values were used to determine if any offset between the two dataset types existed. Any offset found was then applied to the WFDEI data to create values that could be used to gap fill the observational dataset whilst maintaining a consistent mean state.

8.2.5 Model simulations

Land cover fractions were determined from site publications (Table 8.2). For most the tree cover was separate from grass, but broadleaf or needle-leaf were not specified. For the current study it was assumed that trees were broadleaf and that the grass fraction was lawn and hence C3 grass. When the vegetation cover was not sub-divided into trees and grass in the original study, a judgement about the probable vegetation types was made from available satellite imagery.

Values of Q_H and Q_E from the model simulations were used to derive β around mid-day, based on the average of each flux density between 10:00 and 14:00 local solar time (LST) on each

day:

$$\overline{X} = \sum_{i=1}^{N_{days}} \sum_{j=10}^{14} X_{ij} \quad (8.1)$$

where X represents the either Q_H or Q_E, X_{ij} is the flux density at time j ($10{:}00 \leq j \leq 14{:}00$) on day i of the N_{days} of the simulation and $\overline{X}$ is the mid-day average. The long term Bowen Ratio (β) is the ratio of the averaged flux densities:

$$\beta = \frac{\overline{Q_H}}{\overline{Q_E}} \quad (8.2)$$

and was defined the same way for both the observations and model results, with missing observations periods omitted from the calculation. A mid-day value for β was used in preference to a daily average value because both fluxes are likely to be positive and not close to zero, making the Bowen ratio more meaningful.

There could be many sources of errors in the model simulation that could impact on all of the terms within the surface energy balance. However, it is beyond the scope of the current study to investigate the ability of the model to simulate all of these terms. Here the focus is the ability of the model to separate the surface fluxes between turbulent heat and moisture, hence β, and not the individual flux densities.

8.3 Results and discussion

The observed and modelled β values for each of the sites are shown in Fig. 8.1. The model results are in good agreement with observed β at a number of the sites (e.g., B02U, MI95, C95U, see Table 8.1 for sites), but at the majority of the sites β is overestimated by the model. At only one site is β substantially lower than the observed value (VS92). If β is too large, this implies that Q_H is too large compared to Q_E within the model, whilst a value that is too small implies that Q_E is too large compared to Q_H.

In the following discussion we will highlight cases where JULES simulations and measurements disagree and consequently address the issues to explore further possible model improvements.

8.3.1 Influence of garden irrigation

One possible explanation for large β values in the JULES model is that the vegetation could be

soil moisture stressed, and hence the transpiration has been reduced. To investigate the potential impact of the soil moisture, Figure 8.2 shows the initial soil moisture profile within the soil column normalised by the critical point (the point at which vegetation starts to become soil moisture stressed within JULES, Best et al., 2011). Hence a value less than one for any layer indicates that there is reduced soil moisture available to the roots in that layer, which will thus restrict the transpiration accordingly.

Different root density profiles are used within JULES which corresponds to where soil moisture may be removed from by trees and grass (Fig. 8.2). For grass soil moisture can be removed primarily from the second and third soil layers (0.1 - 1.0 m depth) within the model, whilst for tress the third and fourth soil layers (0.35 - 3.0 m depth) are the primary sources.

For many of the sites for which the model over predicts β (TU90, MB03_04, S91U, AR94, SG94, AR93, OU03, Fig. 8.1), the initial soil moisture profile (Fig. 8.2) was below the critical point for at least three of the four soil layers within the model.

To investigate the impact of the soil moisture stress on the vegetation, the model was re-run for each site, but with the unfrozen soil moisture (i.e., the liquid water phase that is available for transpiration) in every layer set to the critical point (or saturation minus frozen soil moisture if this was smaller) at each time-step, i.e., no soil moisture stress for the vegetation. The results of these simulations are shown as β (Fig. 8.3a) and evaporative fraction (i.e., Q_E / (Q_H+Q_E), Fig. 8.3b) to emphasis different parts of the vegetation fraction spectrum. Maintaining the soil moisture at the critical point in each layer reduces β to below that of the observations. Hence the model can represent observed β values, but only if there is no soil moisture stress for the vegetation.

The spin-up strategy (section 2.3) used to initialise the soil moisture for each of the sites should have resulted in a reasonable initial state. The additional soil moisture required to give a good simulation from the model could be the absence of an anthropogenic water injection. This may be undertaken by individuals to maintain their gardens (e.g., S91U) or by the city to clear up after markets (e.g., ME93, MA01). However this may also be regulated; for example alternating days (odd/even) such as S91U (Grimmond and Oke, 2002) or banned such as VS92 because of drought). Under unrestricted irrigation conditions, Q_E has been demonstrated to closely follow irrigation (Grimmond and Oke 1986). The objective of watering gardens is to ensure that the vegetation is healthy, hence it would not be unreasonable to assume that the soil moisture for the majority of patches in an urban neighborhood is maintained around, or above, the critical point during dry periods.

8.3.2 Influence of long-term soil water representation

The simulations that maintained the soil moisture at the critical point also have a beneficial impact in reducing β at the BA01_06 site, and to a much smaller extent for the LO01_02 site, even though the initial soil moisture from the spin-up simulation was above the critical point for these sites.

Both of these sites are multi-year datasets and as such, it is not only the initial soil moisture that will impact on overall β, but also the longer term evolution of the soil moisture during the model simulation. Figure 8.4 shows the initial, final and year end soil moisture profiles for each of the four sites with a data period greater than 12 months. It is clear from Fig. 8.4 that for the BA01_06 site the soil moisture is drying during the simulation, with the bottom layer soil moisture (which has the long term memory) having a dryer soil moisture profile each year than the year before. The same is also true, but to a lesser extent, for the LO01_02 site. The MB03_04 site has almost no change in bottom layer soil moisture, but benefits from setting the soil moisture to the critical point because all of the soil moisture profiles are dry. For the HE07_09 site there is no trend in the bottom layer soil moisture, and hence the soil moisture state is not out of balance. The soil moisture is above the critical point for all of the profiles which is consistent with there being no impact on β when setting the soil moisture profile to the critical point (cf. Fig. 8.1 and Fig. 8.3a).

The trends in soil moisture profile over the period of the simulations for both the BA01_06 and LO01_02 sites could result from relatively dry conditions during the observational period compared to the previous years, or more likely from different mean precipitation values used to force the model between the observations and the WFDEI data used for the spin-up. Whilst it is not possible to compare the average precipitation between these two datasets for the spin-up period, because this is before the observational period, it is possible to compare precipitation during the observational period itself (Fig. 8.5). Fig. 8.5 clearly shows that there is more precipitation in the WFDEI dataset for the BA01_06 site than in the observational dataset. For the other three sites, there is less difference between the average precipitation from WFDEI and the local observations, although for LO01_02 there is slightly more precipitation in WFDEI.

To investigate if an issue within the WFDEI dataset impacts the spin-up, precipitation data were retrieved from synoptic stations close to the four sites: Baltimore Washington International Airport (39.2° N, 76.7° W), Helsinki/Seutula (60.3° N. 25.0° E), Lodz-Lublinek (51.7° N, 19.4°

E) and Melbourne Airport (37.7° S, 144.8° E) (data obtained through NOAA's National Climate Data Center: http://www.ncdc.noaa.gov/). Differences between the synoptic and local observation data, and between the WFDEI and synoptic data are also shown in Fig. 8.5. This clearly shows that for the BA01_06 site, the WFDEI precipitation data are in closer agreement with the local synoptic conditions than the site data, although the synoptic station may have been included in the data analysis used to create the WFDEI dataset. However, the BA01_06 rainfall data have not previously been analysed and detailed analysis of surrounding gauges has not been performed here. For the LO01_02 site the WFDEI dataset are also in better agreement with the synoptic station data, but the differences are much smaller than for the BA01_06 site.

The implications for BA01_06 are that the original simulation using the observed precipitation forcing from the site (Fig. 8.1) had a negative bias in the observations (i.e., too little rainfall). To assess the impact of this, the model was re-run for BA01_06 and LO01_02 with all atmospheric forcing data provided from the WFDEI dataset rather than the local observations. All data were used, rather than just the precipitation data from WFDEI, to ensure consistency between the atmospheric data, e.g., to avoid simulating precipitation from WFDEI under clear sky conditions from the local observations (Fig. 8.3). It has been assumed here that any issues relating to the heights of the various forcing variables from WFDEI when used with JULES can be neglected. Whilst this assumption may not be valid, there are no other options available for obtaining consistent forcing data at more appropriate heights.

The greater precipitation from the WFDEI dataset maintains the soil moisture profile above the critical point for the BA01_06 site and hence β is reduced to values that are less than those observed (evaporative fraction greater than observed), which is consistent with many of the other sites in Fig. 8.3. There is little difference in the resulting β for the LO01_02 site, which is consistent with there being small differences in average precipitation between the WFDEI and local observational datasets.

8.3.3 Influence of bare soil surfaces

Whilst additional anthropogenically applied water might be responsible for maintaining vegetation transpiration rates at many of the sites it is unlikely that unmanaged or bare soil areas are also irrigated. However, in the JULES model the different surface types share the same underlying soil. Hence setting the soil moisture profile to the critical point during the simulation will also unrealistically increase the bare soil evaporation and provide an infinite reservoir of water as conservation of mass is no longer constrained. As both the OU03 and TU90 sites had a

substantial fraction of bare soil or unmanaged land cover, they could be affected by this model limitation.

To investigate the impact on β, Q_H and Q_E, determined by the weighted average of the flux densities from each of the surface types, were reconstructed using the flux densities from the simulation with soil moisture set to the critical point for the vegetation surfaces, and data from the original simulation for the other surfaces. This is equivalent to irrigating only the vegetation part of the land cover.

The resulting β and evaporative fraction values for the TU90 and OU03 sites are shown in Figure 8.3. The higher water availability for bare soil evaporation from the simulation with the soil moisture set to the critical point gave values of β that were substantially lower, and evaporative fractions substantially higher, than those observed (Fig. 8.3). However, irrigating only the vegetated area reduced the unrealistically high β values from the original simulation for these sites (Fig. 8.1), but does not lead to such low values. Indeed, for TU90 the resulting β is in good agreement with the observed values, whist for OU03 the modelled β is higher than observed, but within the range of the observations.

8.3.4 Influence of street cleaning

The modelled β for the ME93 and VL92 sites are substantially larger than the observed values (Fig. 8.1). Setting the soil moisture to the critical point has no impact on modelled β because the fraction of vegetation and bare soil within the footprint is small for both of these sites (1% and 2% for ME93 and 5% and 0% for VL92 respectively). Hence the available water for Q_E must come from a different source to the vegetation or bare soil surfaces.

At the ME93 site, there was daily cleaning of the streets in the morning in preparation for the market (Oke et al., 1999). To understand if this source of water can explain a lower β in the observations at ME93, precipitation was added to the forcing dataset between the local hours of 07:00 and 08:00 each day. In addition, to ensure that the resulting water could only be retained on the street part of the urban surface and not the roofs, the water holding capacity of the roofs was set to zero. The amount of precipitation each day was set to the maximum water holding capacity of the street, which is 0.5 mm in the default parameter settings of JULES (Best et al., 2011). Hence this water reservoir within the street was set to its maximum value at this time, for each day of the simulation. In this scenario, the resulting modelled β is greatly reduced and results in values that are substantially below those observed (Fig. 8.3).

Information of the actual residual water that remained after the street cleaning process is not available from the field study and so it is not clear if the correct amount of water was added to the street surface within the model. A sensitivity study, by varying the amount of artificial daily precipitation, shows that the optimal value of water held within the street to give the same average β as that observed was around 0.2 mm. (Fig. 8.3). Hence it is feasible, and perhaps likely, that the source of water from street cleaning was responsible in reducing β to that observed.

Street cleaning was also undertaken at the MA01 site during mid-morning after the market (Grimmond et al., 2004a). The same artificial total precipitation required to fill the maximum water holding capacity of the street (0.5 mm) and the optimal value obtained for ME93 (0.2 mm) were applied to the MA01 site, expect that the artificial precipitation was added between 10:00 and 11:00 each day. In this case the additional source of water has less of an impact because there is already a Q_E from the irrigated vegetation faction. However β is reduced when the water from street cleaning is added (Fig. 8.3), with 0.2 mm of water having a median that is in better agreement with the observations than 0.5 mm, as for the ME93 site.

8.3.5 Influence of advective fluxes

No additional source of water at the surface was documented during the observational period for the VL92 site. Indeed, during this period Vancouver was experiencing drought conditions and was under an irrigation ban (Grimmond and Oke, 2002). As such, the mid-day Q_E observed are small compared to the net all-wave radiation or the downward component of the short-wave radiation (Fig. 8.6). The small values of Q_E mean that the measurement errors will be larger than normally considered for this term. However, studies of the surface energy imbalance for observational datasets suggest that the error in Q_E usually leads to an underestimation (e.g., Leuning et al., 2012), so any adjustment of Q_E for measurement errors would increase Q_E and thus reduce β. So it is unlikely that the differences between the modelled and observed β values can be explained through observational errors.

As there was no precipitation during the observational period, the only water store at the land surface would be through the soil moisture. Since there was no bare soil surface within the source area of the observed fluxes, this implies that the only possible moisture source from the surface would be through transpiration from the vegetation. Fig. 8.2 shows that the initial soil moisture profile for the VL92 site from the model spin-up was such that there was no soil

moisture stress on the vegetation, even though the site was actually in drought conditions. So the underestimation of Q_E, and hence high β, from the model can not be explained by the initial conditions. This is also confirmed by the run with soil moisture held at the critical point, since this run does not impact on modelled β (Fig. 8.3).

If the observed mid-day values of Q_E at VL92 are scaled by the vegetation fraction, f_v (i.e., the only water vapour originates from transpiration from the vegetation fraction of the land cover), then the resulting Q_E is larger than the net all-wave radiation (Fig. 8.6) and even larger than the downward component of the shortwave radiation at times. Hence it is very unlikely that Q_E observed at the VL92 site originated from soil moisture through transpiration within the turbulent source area of the eddy-covariance observations on the tower.

As this was an industrial site, although there was no street cleaning documented, it is possible that there were some equivalent activities that could lead to a source of water on impervious surfaces. As such, a simulation with sufficient artificial total precipitation to fill the maximum water holding capacity of the street (0.5 mm) was applied for each hour between the working hours of 09:00 and 17:00 on each day. The addition of this water each hour provides a source reservoir that is large enough to reduce β to values far less than observed (Fig. 3). However, a sensitivity study shows that an amount of 0.035 mm each hour gives a modelled median of β that is close to that observed (Fig. 3). Hence it requires only a small amount of water to be added at each hour to explain the observed β, so it is possible that such a source of water is responsible for the observed evaporation.

An alternative explanation is that the moisture originates from the advective flux at atmospheric levels below the height of the eddy covariance system. Indeed, the wind direction around mid-day for most of the observational period was from the direction of False Creek, an inlet of the Pacific Ocean located 600 m to 1 km upwind of the tower. A relatively warm and dry surface such as that within the observational footprint could give the buoyancy required to lift the advected vapour flux at low levels, hence leading to an observed mid-day average Q_E of 36 W m^{-2} at the site.

8.3.6 Influence of a garden irrigation ban

The VS92 site is the only site where the model substantially underestimates observed β. The observational period for this site coincided with the VL92 site, so was also experiencing drought conditions with an irrigation ban. However, the initial soil moisture profile for the model

derived from the spin-up has a soil moisture profile that is above the critical point, and hence the vegetation in the model is not soil moisture stressed (Fig. 8.2). This implies that the there was too much precipitation in the forcing data from the WFDEI dataset during the spin-up period, especially during the period immediately prior to the start of the observations.

Observations for the VS92 dataset were taken over 56 days, during which there was no precipitation in either the observational dataset or the WFDEI dataset. Therefore it is not possible to make conclusions about any biases that there could be in the WFDEI dataset compared to the observations. In addition, the complex topography of the Vancouver area and the proximity of the site being a coastal location results in large precipitation gradients across the city (Oke and Hay, 1998). As such, comparing the WFDEI dataset to a synoptic station would not necessarily result in any conclusions about precipitation biases compared to the observational site. Moreover the WFDEI dataset has a resolution of 0.5^{o} and as such can not be expected to give accurate precipitation values for specific parts in a region of such topographic heterogeneity.

The WFDEI dataset has two precipitation datasets based upon monthly climatologies from either GCPC or CRU (section 2.3). In this study we have used the values from the GCPC data, but both climatologies are based upon a similar global precipitation gauge network. The number of gauges used for the climatology has a much lower density in the Vancouver (Canada) region compared to the coastal regions just to the south in the USA (see Schneider et al., 2013, Fig. 8.5). Also, New et al. (2000, their Fig. 8.1) show that the rain gauge density used for the CRU climatology decreased substantially between 1981 and 1995. Hence it is quite likely that with the heterogeneous nature of precipitation around Vancouver, the rain gauge density during the period of the observational campaign could have resulted in a lower quality precipitation product for this site compared to other regions that have higher gauge densities. Thus the 10 year spin-up for both VL92 and VS92 could be impacted.

Irrigation restrictions were also enforced during the summer at the MB03_04 site. However unlike the complete ban at VS92 and VL92, at MB03_04 this involved no watering of lawns, whilst for trees and other vegetation automatic sprinkler systems were limited to the hours of 23:00 and 06:00, and manual sprinkler systems limited to the hours of 05:00 - 08:00 and 20:00 - 23:00. In addition, although the times during which irrigation could be applied were limited, the amount of water was not.

Calculating an average β for both the summer and winter at the MB03_04 site shows that although β is slightly reduced in the winter, there is no impact on the ability of the model to

simulate the observed values if it is assumed that the vegetation is sufficiently irrigated (not shown). The summer values for both observed and modelled β are similar to the overall results. Hence the partial irrigation ban for the MB03_04 site has little impact on the overall β compared to the complete ban at the VS92 site.

8.4. Conclusions

The initial soil moisture conditions have been shown previously to be critical for modelling sensible and latent heat fluxes in urban environments (Best and Grimmond, 2015). In this study, initialising soil moisture with saturated conditions prior to a 10 year spin-up is shown to produce a soil moisture profile that is consistent with the model physics whilst enabling a realistic simulation. Hence we recommend this for future studies when soil moisture profile observations are unavailable.

In addition, the WFDEI dataset is demonstrated to in general provide good quality forcing data that can be used with this spin-up strategy. Whilst the quality of the precipitation data within the WFDEI dataset can vary depending upon the rain gauge density used to create monthly climatologies such as GPCC and CRU, it was of sufficient quality for most of the sites considered in this study. Hence we also conclude that by using the WFDEI data and the 10 year spin-up strategy, it should be possible to initialise a LSM (including ULSM) at any site, as long as consideration is given to the density of rain gauges used for the monthly precipitation climatology.

In a summary of the results from PILPS-Urban, Best and Grimmond (2015) concluded that the important processes in the urban environment were the bulk reflection of the downward shortwave radiation, the influence of the urban morphology on the longwave radiation fluxes and the vegetation processes for the distribution of the sensible and latent heat fluxes. This study has focused on the ability of JULES to simulate β across 22 observational datasets, i.e., exploration of the model's ability to partition surface energy between the sensible and latent heat fluxes, hence the third physical process identified by Best and Grimmond (2015) is addressed. However, a good simulation of β does not necessarily imply that the model gives accurate values of Q_H and Q_E separately, which are also influenced by the radiative processes.

The results from the model show that at sites where the transpiration from vegetation is not restricted by limited soil moisture the model can reproduce observed β, whilst for the sites with limited soil moisture the model overestimates β compared to the observations. However, if we

make the assumption that urban sites are irrigated to ensure that vegetation is not soil moisture stressed (i.e., urban residents maintain 'healthy' gardens and parks), then the model is in good agreement with observed β at these sites as well. The one exception, the VS92 site, was known to be in drought conditions with an irrigation ban in force. Hence we conclude that when modelling vegetation within urban environments it should be assumed that the vegetation is not soil moisture stressed, unless it is known to be a dry period with an irrigation ban in place.

The possibility of an irrigation ban within urban environments makes the modelling of urban vegetation complex, but important. The availability of soil moisture for transpiration is not a physical condition as it is for the rural environment, but becomes a combination of physical and social conditions. Factors such as population density (i.e., water demand), wealth (e.g., artificial water storage), national infrastructure (i.e., transport of water) and stake holder requirements (e.g., city dweller water use versus agricultural irrigation) may all contribute towards the political decision making with regards to an irrigation ban. For instance, compare the different urban water use practices and water availability in the climates of Ouagadougou (Offerle et al., 2005b), Marseille (Grimmond et al., 2004a), Vancouver and Chicago (Grimmond and Oke, 1999a), and Arcadia and San Gabriel (Grimmond et al., 1996). Hence we conclude that there should be further studies investigating the implementation of irrigation bans and their impact on the surface energy balance for urban areas.

Irrigation of vegetation is not the only anthropogenic moisture source that can influence the turbulent fluxes of heat and moisture within the urban environment. This work has shown that maintenance activities such as street cleaning can provide a source of water that can moderately increase Q_E. Hence all possible sources of anthropogenic water availability are important and need to be represented within an urban land surface model. Furthermore, the impact of such anthropogenic water injections suggests that they are at least as important as the anthropogenic heat flux density on the terms in the surface energy balance for urban areas.

For well irrigated vegetation, there is little change in β for sites with vegetation cover between 20-30% and 70-90%. Whilst there is some day to day variability at the sites, the average β is typically in the range of 1-2. The two rural sites with almost total vegetation cover (B02R1, B02R2) have β values less than one, which is typical for rural locations. However, as the vegetation fraction decreases below 20-30%, β increases substantially, with a maximum value of around eight for the most densely built up urban site studied here (ME93). However, for this site β was reduced due to being controlled by water availability from street cleaning. Also, for the second most impervious site (VL92) the observations may have been influenced by water added to the surface in a similar manor to street cleaning, or atmospheric advection of moisture

into the source area at levels below the height of the observations. As such, it is possible that without these additional sources of moisture, β could be as large as 20 for urban sites with little vegetation during summertime. However, Offerle et al. (2006b) suggested that sparse vegetation may well be exposed to higher vapour pressure deficits and higher temperatures, whilst isolated trees are exposed to higher Photosynthetic Active Radiation (PAR), which could increase transpiration. Also, Meier and Scherer (2012) concluded that trees surrounded by a high fraction of impervious surfaces showed consistently higher canopy temperatures. In addition, we have made no attempt to distinguish between native and non-native vegetation. High latitude, mid-latitude, semi arid and tropical vegetation all have different characteristics which could influence the results for sparse vegetation cover. Hence additional observational studies are required for urban environments with sparse vegetation and no additional anthropogenic water injections,to determine the behavior of vegetation in such environments.

If we consider how Q_E varies with vegetation fraction, we find that as a proportion of the available energy at the surface, there is a step change around vegetation fractions of 20-30% (Fig. 8.7a). This step change is also seen when scaling Q_E by the incoming all-wave radiation (Fig. 8.7b). This result agrees with Loridan and Grimmond (2012a) who found such a step change in the scaled Q_E against their active vegetation index. Furthermore, when scaled by the incoming all-wave radiation, there is also a step change in Q_E with almost total vegetation cover (70-90%, Fig 8.7b), or little built area cover, although this step change is not seen in Q_E as a proportion of the available energy at the surface. This suggests that there could be a step change in the net heat storage flux density for small built fractions, as confirmed by the results of Loridan and Grimmond (2012a) who showed a step change in the storage heat density for changes in active built index. Hence we conclude that the sensitivity of Q_E, and hence the Q_H through the available energy at the surface, is greatly increased when there is little vegetation cover, whilst the sensitivity of the heat storage is greatly increased when there is little built area cover.

The results from this study suggest that an urban land surface model, such as JULES, can reproduce the observed β values of urban sites. However, the sensitivity of the urban energy balance at sites with low fractions of vegetation land cover, or low fractions of built area, suggests that further studies are required for urban environments with less than 30% vegetation cover, and less than 30% built area cover. This can only be achieved if there are future observational campaigns for such environments, or observational data are analysed according to wind sectors that have differing plan area vegetation fractions. There is a need for future observational campaigns to be long term in order to sample a range of synoptic and climatic extremes, so that the nature of the variability and sequencing can be evaluated for their impacts

on the surface fluxes.

Table 8.1: Sources of observational data used in the analyses, with the main references for the data and the site characteristics.

City	Site	Lat / Long	Period / Length	Instrument height (m)	References
Arcadia, CA, USA	AR93	34.1 N 118.0 W	Jul - Aug 93 40 days	30.5	Grimmond and Oke, 1995 Grimmond and Oke, 2002
Arcadia, CA, USA	AR94	34.1 N 118.0 W	Jul 94 19 days	32.8	Grimmond et al., 1996 Grimmond and Oke, 2002
Baltimore	BA01_06	39.2 N 76.7 W	May 01 – Dec 06 2049 days	41.2	Crawford et al., 2011
Basel, Switzerland	B02R1	47.5 N 7.7 E	Jun – Jul 02 30 days	1.5 – 28.0	Christen and Vogt, 2004
Basel, Switzerland	B02R2	47.6 N 7.6 E	Jun – Jul 02 30 days	2.0 – 3.3	Christen and Vogt, 2004
Basel, Switzerland	B02S1	47.6 N 7.6 E	Jun – Jul 02 30 days	15.0 – 15.8	Christen and Vogt, 2004
Basel, Switzerland	B02U1	47.6 N 7.6 E	Jun – Jul 02 30 days	25.5 – 31.7	Christen and Vogt, 2004
Basel, Switzerland	B02U2	47.6 N 7.6 E	Jun – Jul 02 30 days	33.0 – 37.6	Christen and Vogt, 2004
Chicago, IL, USA	CH92	41.6 N 87.5 W	Jul 92 13 days	18.0	Grimmond and Oke, 1995, 2002
Chicago, IL, USA	C95U	41.6 N 87.5 W	Jun - Aug 95 57 days	27.0	King and Grimmond, 1997 Grimmond and Oke, 2002
Helsinki, Finland	HE07_09	60.2 N 24.9 E	Jan 08 - Dec 09 853 days	31.0	Vesala et a., 2008 Jarvi et al., 2014
Lodz, Poland	LO01_02	51.8 N 19.5 E	Mar 01 - Dec 02 730 days	37.0	Offerle et al., 2005a, 2006a,b Pawlak et al., 2011
Marseille, France	MA01	43.2 N 5.2 E	Jun - Jul 01 27 days	39.0	Grimmond et al., 2004a
Melbourne, Australia	MB04_05	37.8 S 144.9 E	Aug 04 – Nov 05 475 days	35.0	Coutts et al., 2007a Coutts et al, 2007b
Mexico City, Mexico	ME93	19.2 N 99.1 E	Dec 93 7 days	28.0	Oke et al., 1999 Grimmond and Oke, 2002
Miama, FL, USA	MI95	25.4 N 80.2 W	May - Jun 95 26 days	40.8	Newton, 1999 Grimmond and Oke, 2002 Newton et al., 2007
Ouagadougou, Burkina Faso	OU03	12.2 N 1.3 E	Feb 03 26 days	10.0	Offerle et al., 2005b
San Gabriel, CA, USA	SG94	34.1 N 118.0 W	Jul 94 22 days	18.0	Grimmond et al., 1996 Grimmond and Oke, 2002
Sacramento, CA, USA	S91U	38.3 N 121.3 W	Aug 91 10 days	29.0	Grimmond et al., 1993, Grimmond and Oke, 1995, 1999a, 2002
Tuscon, AZ, USA	TU90	32.1 N 110.6 W	May – Jun 90 45 days	25.6	Grimmond and Oke, 1995, 2002
Vancouver, Canada	VL92	49.3 N 123.1 W	Aug 92 15 days	9.0	Grimmond and Oke, 1999a, 2002
Vancouver, Canada	VS92	49.3 N 123.1 W	Jul – Sep 92 56 days	14.1	Grimmond and Oke, 1999a, 2002

Table 8.2: Sites ordered by increasing plan vegetation area within the footprint of the observations. See Table 8.1 for site names and sources of data. * indicates the sites where judgement had to be used to determine tree and grass cover.

Site	Trees (%)	Grass (%)	**Total Veg (%)**	Buildings (%)	Streets (%)	**Total built (%)**	Bare soil (%)	Water (%)
ME93	1	0	1	54	44	97	2	0
VL92	3	2	5	51	44	95	0	0
OU03	10	0	10	40	20	60	30	0
MA01	13	0	13	60	27	86	1	0
B02U1	0	16	16	54	30	84	0	0
TU90	11	7	18	23	42	65	17	0
B02U2	0	31	31	37	32	69	0	0
LO01_02	31 *	0 *	31	30	40	69	0	0
MI95	7	27	34	35	29	64	0	2
SG94	12	25	37	29	31	60	0	4
MB04_05	23	15	38	45	18	62	1	0
C95U	7	32	39	36	25	61	0	0
VS92	9	35	44	31	24	55	2	0
CH92	10	34	44	33	22	55	1	0
HE07_09	24	25	49	15	36	51	0	0
S91U	13	34	47	36	12	48	1	5
B02S1	0	53	53	28	19	47	0	0
AR94	30	23	53	24	19	43	2	2
AR93	32	24	56	22	18	40	2	2
BA01_06	54	14	67	16	15	31	1	1
B02R1	0	91	91	2	7	9	0	0
B02R2	0	98	98	0	2	2	0	0

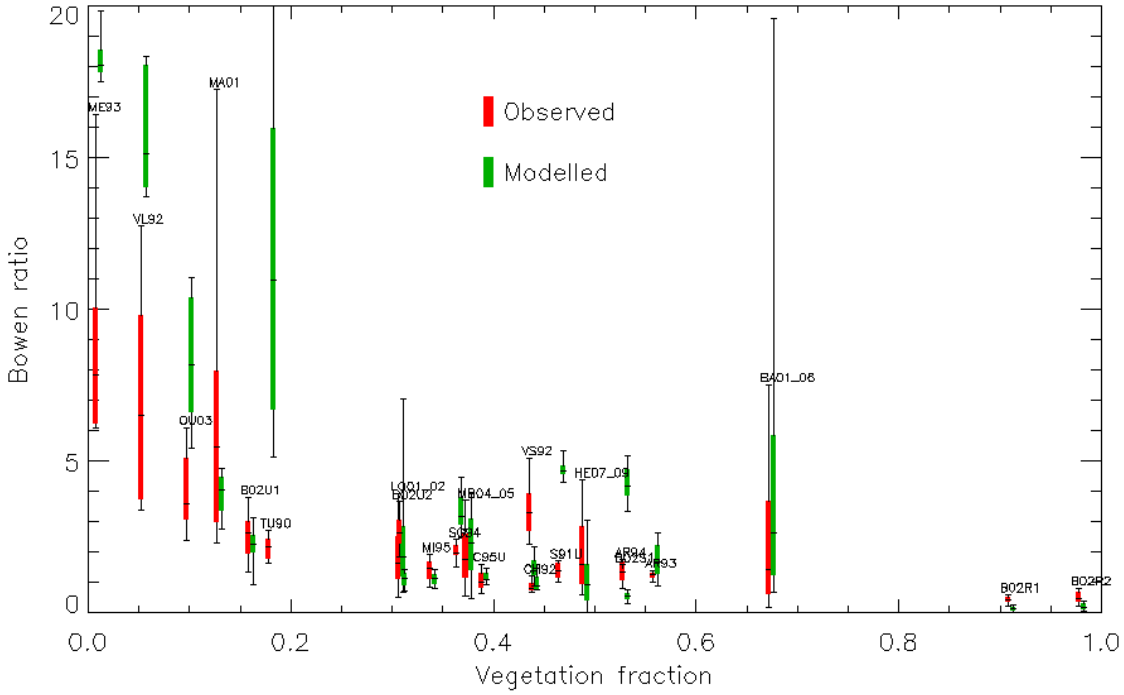

Figure 8.1: Mid-day (10 - 14 local solar time) variability of observed and modelled Bowen ratio shown by the inter-quartile range (box), median (-) and 10th and 90th percentiles (whiskers). Codes for the sites are given in Table 8.1.

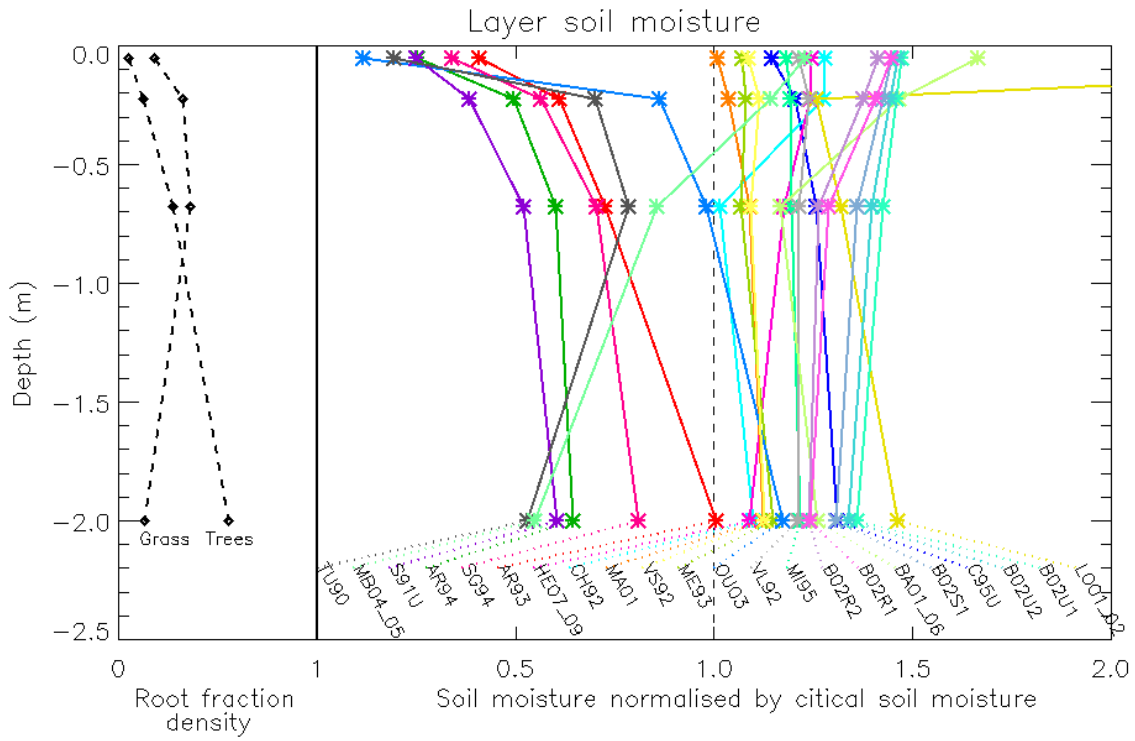

Figure 8.2: Initial soil moisture profile used in the model simulations at each site (Table 8.1) derived from the spin-up, and the model root density profiles for a grass and tree land cover type.

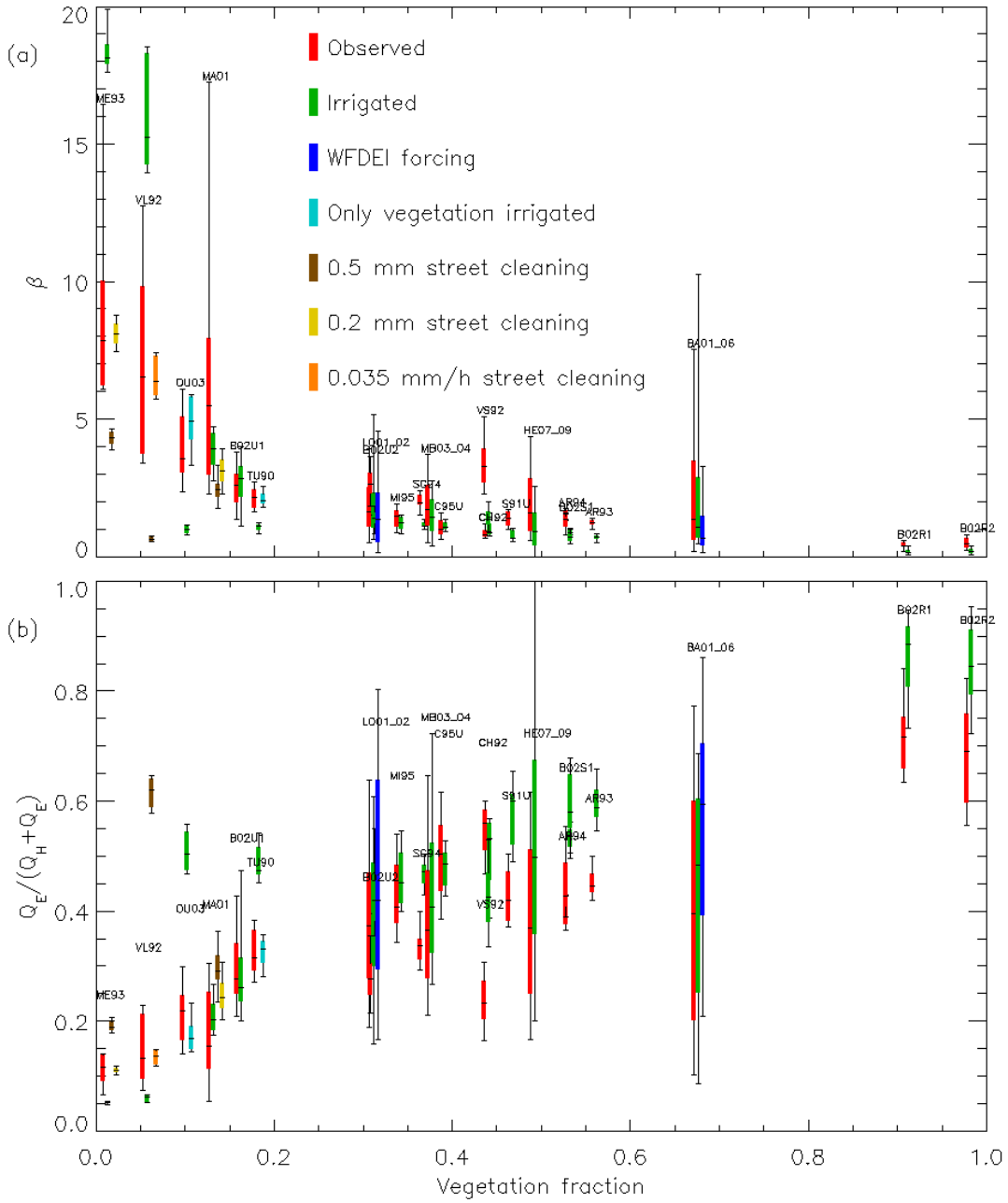

Figure 8.3: (a) as for Fig.8.1 except that anthropogenic moisture has been added in various ways: Irrigated – soil moisture fixed at the critical point; WFDEI forcing – WFDEI precipitation instead of local observations; only vegetation irrigated – soil moisture held at critical point for vegetated land cover only; 0.5 mm street cleaning – precipitation added to forcing in the morning amounting to a total of 0.5 mm; 0.2 mm street cleaning – precipitation added to forcing in the morning amounting to a total of 0.2 mm; 0.035 mm/h street cleaning – precipitation added to forcing each hour between 09:00 – 17:00 amounting to a total of 0.035 mm every hour. See text for further discussion. (b) As for (a) but for evaporative fraction.

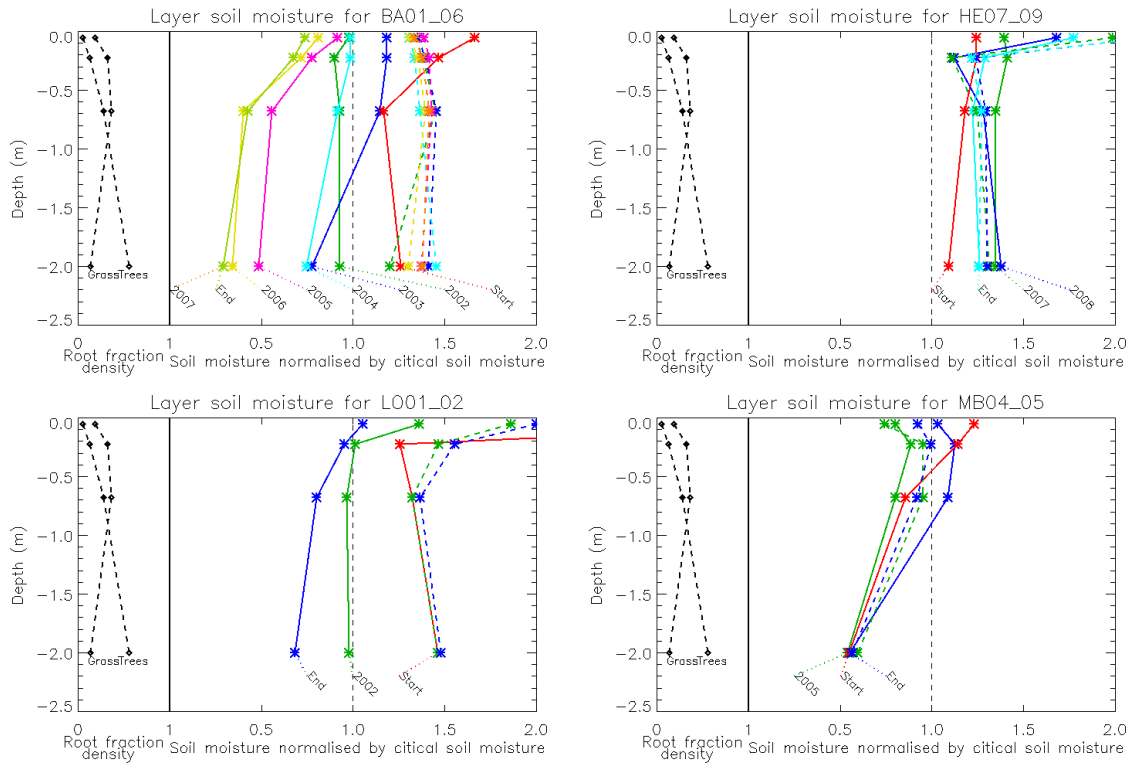

Figure 8.4: As for Fig. 8.2 but for initial, final and end of calendar year soil moisture profiles for model integrations at multi-year observational sites.

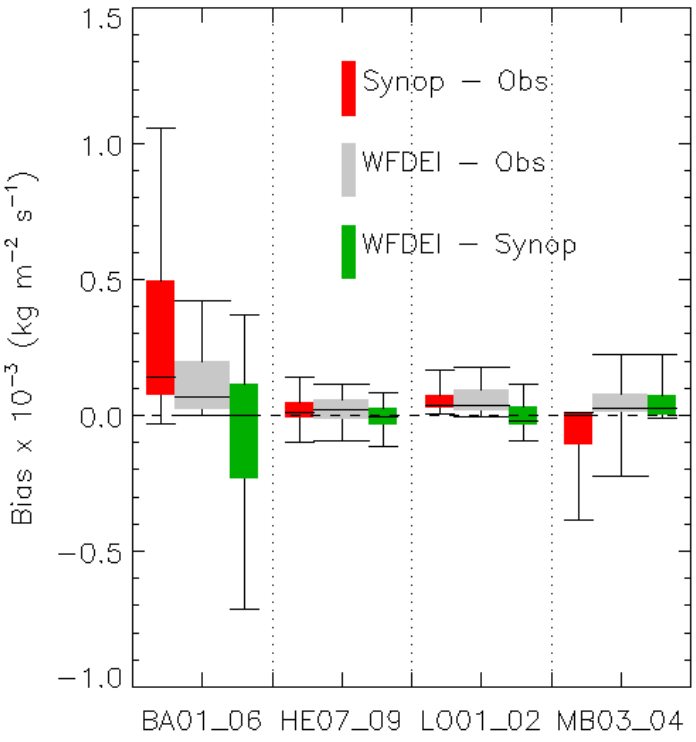

Figure 8.5: Distribution (inter-quartile range, median and 10th and 90th percentiles, as per Fig. 8.1) of bias in precipitation between WFDEI and local observations, synoptic observations and local observations, and WFDEI and synoptic observations for the multi-year sites.

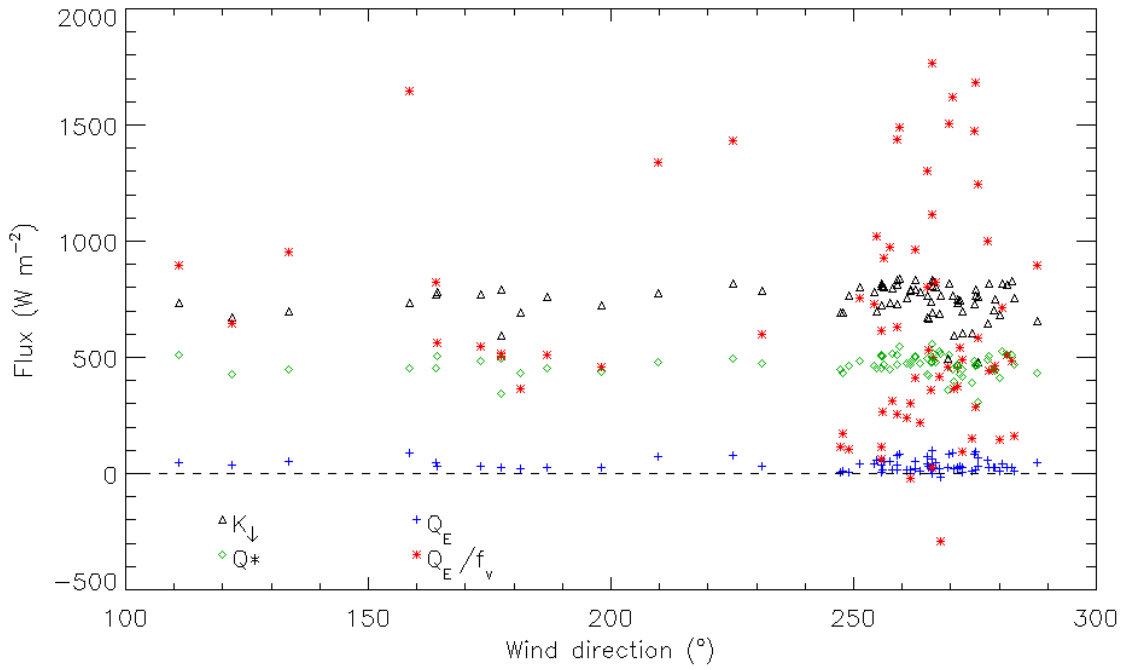

Figure 8.6: Mid-day (10 – 14 local solar time) downward solar radiation (K|), net all-wave radiation (Q*), latent heat flux (QE) and latent heat flux scaled by fraction of vegetation land cover within observational footprint (QE/fv), against wind direction for the VL92 site.

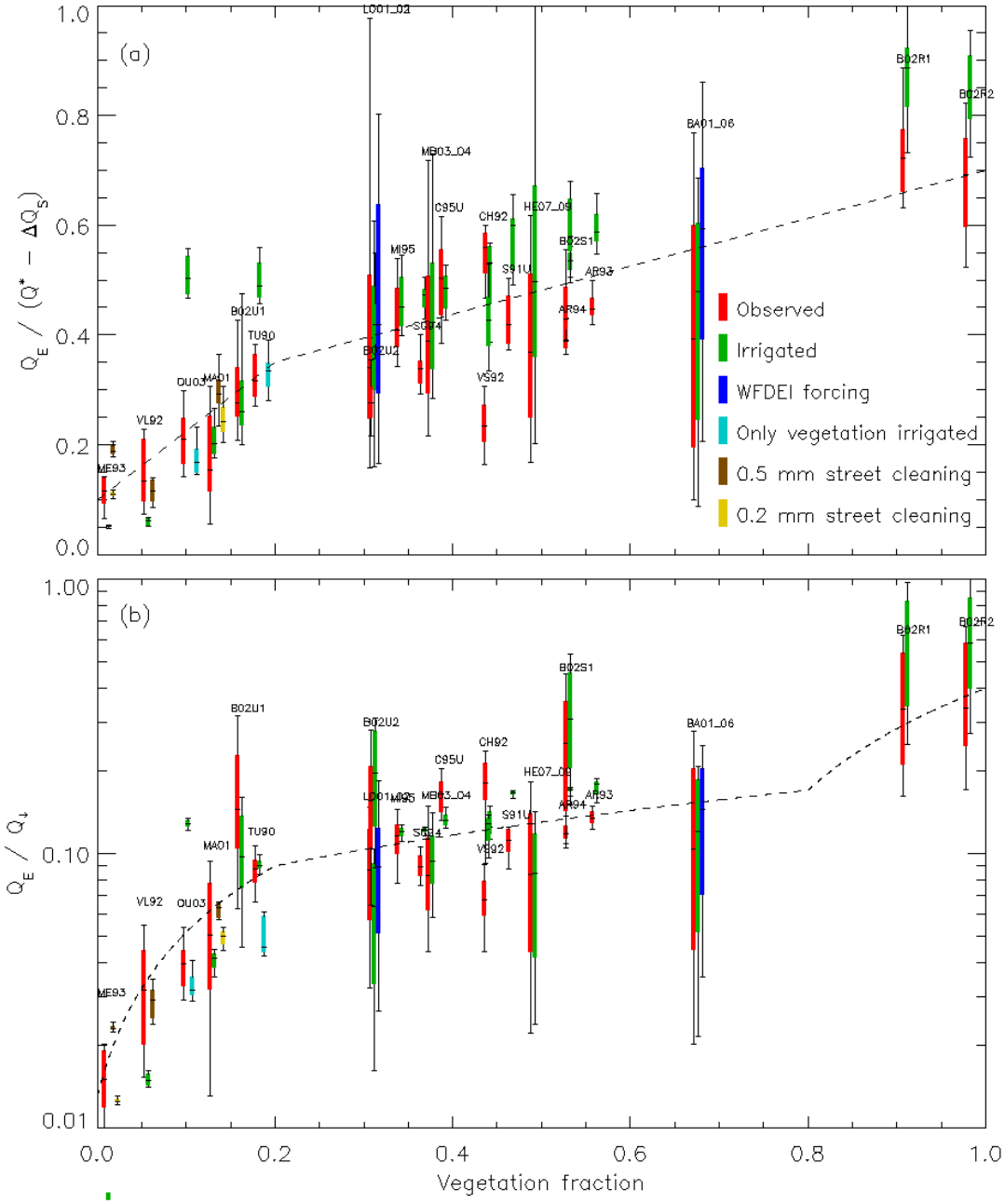

Figure 8.7: Mid-day (10 - 14 local solar time) variability of observed and modelled (a) latent heat scaled by available energy and (b) latent heat scaled by incoming all-wave radiation, shown by the inter-quartile range (box), median (-) and 10th and 90th percentiles (whiskers). Codes for the sites are given in Table 8.1. Linear dashed lines have no significance and are purely a visual guide.

Chapter 9: Conclusions and recommendations

The objectives of this PhD were to establish the level of complexity required in an urban surface exchange scheme for the application of weather forecasting and climate modelling (i.e., accurate prediction of the turbulent sensible and latent heat flux exchange with the atmospheric boundary layer), identify the dominant physical processes within an urban environment for such applications, assess the seasonal behaviour of the urban models and the influence of initial conditions, understand the impact of the anthropogenic heat flux and assess if an urban scheme can represent the surface fluxes over a wide range of vegetation fractions in land cover.

This research has analysed the results for an urban model comparison, from which the following conclusions can be drawn:

The dominant physical processes impacting on the urban surface energy balance (with a focus on the surface/atmosphere exchange of sensible and latent heat fluxes) in order are:

1. The albedos of building materials and shortwave trapping from canyon geometry that influence the amount of shortwave radiation absorbed by the urban surface.
2. A reduced sky view factor from the urban geometry combined with the non-linear combination of outgoing longwave radiation from roofs and canyons, impacting on the net longwave radiation.
3. Evaporation from vegetation which influences the Bowen ratio.

Hence to accurately predict the sensible and latent heat flux exchange between the urban surface and the atmospheric boundary layer, an urban scheme must be able to utilise the bulk surface albedo, be able to distinguish between roofs of buildings and urban canyons, and have a representation of vegetation. To achieve this an urban model requires parameter information of the bulk surface albedo, the height to width ratio of the urban canyons and the fraction of building roofs to the urban canyons, and the vegetation fraction. This information can be determined globally from satellite data and/or simple formulations utilising satellite data.

In general urban models are able to capture the seasonal cycle of the surface fluxes in the energy balance, although they tend to have larger errors in the summer than in the winter, as might be anticipated as the fluxes themselves are larger. On average, only the sensible heat flux has a positive bias throughout the year, with the other surface fluxes (net all wave radiation, latent and net storage heat fluxes) having a negative bias. Moreover, in general the models perform best when there is no time shift between the model results and the observational data, implying that

the models do not have an overall phase error. A simple representation of the physical processes performs at least as well as more complexity, apart for the representation of vegetation. Not enabling a source of moisture for evaporation through vegetation or other processes results in no model data points for the latent heat flux being within 100% of the observed flux.

Due to the importance of representing vegetation within an urban area, the initial conditions of soil moisture are critical. If the soil moisture is set too dry the model fluxes will be impacted until there is sufficient rainfall, whereas if set too wet then the model is not able to restrict evaporation as observed. The impact of incorrect initialisation of soil moisture can influence the model fluxes for over a year and affects not only the latent heat flux but also the sensible and net storage heat fluxes.

The model errors caused by initial soil moisture have implications for future observational studies in urban environments. To make maximum benefit of these observations, an estimate of the soil moisture profile, or at least an indication of the soil moisture state, is required. However, if this is not available then a spin-up strategy using atmospheric forcing data from the WFDEI dataset can give a good estimate of the initial soil moisture, if a 10 year spin-up period is undertaken. Good initial conditions from this strategy are not guaranteed, however, due to assumptions in the algorithms and source datasets used to create the WFDEI forcing, so care should be taken to understand the limitation of this approach.

Urban areas can have a high thermal inertia due to the construction materials used in the buildings. However, the initial condition of the surface temperatures of the various urban facets is not as important as for that of soil moisture. The spin-up timescale for the surface temperatures is less than that for soil moisture and the impact of initial temperature errors are not seen after a seasonal cycle.

The majority of models that contributed to the urban model comparison are generally able to capture the observed trends in the surface fluxes from each of the atmospheric forcing variables. This includes the non-linear trends with respect to humidity and wind speed. However, the models are not able to capture the increasing trend in the net storage heat flux with increasing downward shortwave radiation. Overall the models put too much available energy at the surface into the sensible heat flux at the expense of the net storage heat flux. Improving the models with respect to these biases is not straight forward due to the interaction of each flux with the surface temperature and the implications of thermal properties on the phase and magnitude of the fluxes.

For the models in the comparison that were not able to capture the observed trends in the surface fluxes with regards to the atmospheric forcing, these models do not share common characteristics in their representation of the physical processes, so no general conclusions can be drawn in this regard. More likely it is the way in which the processes have been implemented within the parameterisations rather than the processes themselves.

Conclusions that models which neglect the anthropogenic heat flux performed at least as well as those that include this flux, determined from initial studies of the results from the urban model comparison, may have been influenced by the representation of vegetation within the models. Results from one of the models that contributed to the comparison demonstrate that additional benefit can be obtained from including the anthropogenic heat flux, but the benefit is small due to the size of the anthropogenic heat flux itself. For the Melbourne data, having a temporal variation in the anthropogenic heat flux gives no additional benefit over including a constant value. It is not only the size of the flux that determines this result, but also the fact that the times in the diurnal cycle at which the surface energy balance is likely to be at its most sensitive to this additional source of energy, are the times when the flux itself has its smallest values.

Despite the positive impact of the anthropogenic heat flux being small at the Melbourne site, it is anticipated that the benefits would be much greater at sites with larger anthropogenic heat fluxes. In addition, a larger magnitude in the diurnal variation of the anthropogenic heat flux is also likely to demonstrate additional benefits from a temporal representation of the flux. However, for this to be demonstrated observational campaigns in downtown areas with large anthropogenic heat emissions are required.

In addition to anthropogenic activities leading to additional heating in urban environments, additional water sources through irrigation and street cleaning can also play a critical role in the surface energy balance. For most studies it can be assumed that vegetation is never soil moisture stressed, due to anthropogenic irrigation of gardens and parks, unless restrictions have been applied through complete irrigation bans.

Over most of the range of plan area vegetation cover (30% - 70%), there is a similar response to the distribution of available energy at the surface into sensible and latent heat, with Bowen ratio values around 1 - 2. However, for low vegetation cover (< 20%) there is an increased sensitivity in the Bowen ratio to the vegetation cover, partly due to the low values in the denominator of this parameter, with values potentially as high as 20. For high vegetation cover (> 90%) there is a suggestion that there is sensitivity in the net storage heat flux which impacts on the fraction of the downward net radiation that results in latent heat.

Despite the suggestions of how the surface energy balance terms might be sensitive to the plan area fraction of vegetation, there are observational dataset gaps in the continuum of vegetation fraction, in particular, between 20% - 30% and 70% - 90%. As such, further observational campaigns are needed to fill these gaps and to provide a better understanding of how the surface energy balance components are sensitive to vegetation cover. However, there is no guarantee that such sites can be identified or observational campaigns undertaken. Hence an alternative could be to study the performance of models for differing wind directions that are likely to have different plan area vegetation fractions within their source areas.

The key scientific contributions of the research within this thesis include:

a) Identifying the dominant physical processes within an urban environment that are important for accurate prediction of the turbulent sensible and latent heat flux exchange with the atmospheric boundary layer;
b) Describing the level of complexity of an urban surface energy balance model for this application;
c) Establishing the parameter requirements for such a model;
d) The first consistent evaluation of many urban models over a full seasonal cycle;
e) Identifying the critical impact of initial soil moisture on urban model performance, given the importance of vegetation;
f) The recommendation that future urban observational campaigns should include an assessment of soil moisture;
g) The demonstration that in the absence of a soil moisture estimate, that a spin-up strategy involving the WFDEI data is recommended;
h) Anthropogenic irrigation can have as significant an impact on the surface energy balance fluxes as the anthropogenic heat flux ;
i) It can be assumed that urban vegetation is not soil moisture stressed unless a specific irrigation ban is enforced;
j) The JULES land surface model is able to represent the partitioning of turbulent sensible and latent heat fluxes within urban environments spanning a wide range of vegetation fraction coverage.

The conclusions from the first international urban model comparison are only for the two sites. To ensure that these conclusions are robust, additional sites should be considered. Ideally a more extensive urban model comparison study could be undertaken using a wider range of urban sites (e.g., Chapter 8 and other data sets). This would allow a wider range of morphologies, building materials, vegetation fractions, cultural habits and meteorological

conditions to be assessed. The modelling community should support the observational data sets being established and maintained. This would permit more rapid analysis of these data for comparisons.

Alternative urban model applications may have different model complexity requirements due to their nature (e.g., dispersion within a street canyon). However, the commonality of complexity required for various applications is an important scientific question that needs to be addressed. Hence, a suitably comprehensive urban dataset should be identified (or collected) to assess urban models over a variety of applications.

The organisation of such community comparisons is complex and difficult to achieve in practice. The first urban model comparison experiment was conducted under the banner of PILPS (Project for Intercomparison of Land-surface Parameterization Schemes), a GEWEX (Global Energy and Water cycle EXchanges project) GLASS (Global Land Atmosphere System Study) panel activity. Maintaining and expanding linkages between these two communities to share knowledge and practices will be mutually beneficial, helping to enable future model comparisons and integrate urban areas within wider land modelling activities.

References

Ackerman TP (1977) A model of the effect of aerosols on urban climates with particular applications to the Los Angeles Basin. J. Atmos. Sci. 34, 531-547.

Allen L, Lindberg F, Grimmond CSB (2011) Global to city scale urban anthropogenic heat flux: model and variability. Int J Climatol 31: 1990-2005. doi: 10.1002/joc.2210

Anandakumar K (1999) A study on the partition of net radiation into heat fluxes on a dry asphalt surface. Atmos Environ 33: 3911–3918.

Avissar R, Pielke RA (1989) A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. Mon Weather Rev 117: 2113-2136

Barlow JF, Harman IN, Belcher SE (2004) Scalar fluxes from urban street canyons. Part I: Laboratory simulation. Boundary-Layer Meteorol. 113: 369–385.

Best, MJ (2005) Representing urban areas within operational numerical weather prediction models. Boundary-Layer Meteorol, 114: 91–109.

Best MJ (2006) Progress towards better weather forecasts for city dwellers: From short range to climate change. Theor Appl Climatol 84: 47-55

Best, M.J., Grimmond CSB (2013) Analysis of the seasonal cycle within the first international urban land surface model comparison, Boundary-Layer Meteorol., 146, 421-446, doi: 10.1007/s10546-012-9769-7.

Best MJ, Grimmond CSB (2014) Importance of initial state and atmospheric conditions for urban land surface models performance. Urban Climate, 10, 387-406, doi:10.1016/j.uclim.2013.10.006

Best MJ, Grimmond CSB (2015) Key conclusions of the first international urban land surface model comparison. Bulletin of the American Meteorol. Soc. 96, 805-819. doi:http://dx.doi.org/10.1175/BAMS-D-14-00122.1

Best MJ, Grimmond CSB, Villani MG (2006) Evaluation of the urban tile in MOSES using surface energy balance observations. Boundary- Layer Meteorol, 118: 503–525.

Best M.J., Pryor M , Clark DB, Rooney GG, Essery RHL, Ménard CB, Edwards JM, Hendry MA, Porson A, Gedney N, Mercado LM, Sitch S, Blyth E, Boucher O, Cox PM, Grimmond CSB, Harding RH (2011) The Joint UK Land Environment Simulator (JULES), Model description – Part 1: Energy and water fluxes. Geosci Model Dev, 4: 677-699

Best, MJ, Abramowitz G, Johnson HR, Pitman AJ, Balsamo G, Boone A, Cuntz M, Decharme B, Dirmeyer PA, Dong J, Ek M, Guo Z, Haverd V, van den Hurk BJJ, Nearing GS, Pak B, Peters-Lidard C, Santanello Jr. JA, Stevens L, Vuichard N (2015) The plumbing of land surface models: benchmarking model performance. J. Hydrometeorol. 16, 1425-1442,

doi:10.1175/JHM-D-14-0158.1.

Blankenstein S, Kuttler W (2004) Impact of street geometry on downward longwave radiation and air temperature in an urban environment. Meteorologische Zeitschrift 13, 373-379.

Bohnenstengel, SI, Evans S, Clark PA, Belcher SE (2011) Simulations of the London urban heat island, Q. J. R. Meteorol. Soc., 137, 1625-1640, doi: 10.1002/qj.855.

Bornstein, R, Lin Q (2000) Urban heat islands and summertime convection thunderstorms in Atlanta: three case studies, Atmos. Environ., 34, 507-516, doi:10.1016/S1352-2310(99)00374-X.

Brunner, D, Lemoire G, Bruzzone L, Greidonus H (2010) Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique. IEEE Transactions on Geoscience and Remote Sensing, 48, No.3, doi:10.1109/TGRS.2009.2031910.

Carlson TN, Boland FE (1978) Analysis of urban-rural canopy using a surface heat flux/temperature model. J. Appl. Meteorol. 17, 998-1013.

Chaxel, E., Chollet J-P (2009) Ozone production from Grenoble city during the August 2003 heat wave, Atmos. Environ., 43, 4784-4792, doi:10.1016/j.atmosenv.2008.10.054.

Chen F., Kusaka H, Tewari M, Bao J, Hirakuchi H (2004) Utilizing the coupled WRF/LSM/Urban modeling system with detailed urban classification to simulate the urban heat island phenomena over the Greater Houston area. Fifth Symposium on the Urban Environment, CD-ROM. 9.11. Amer. Meteor. Soc., Vancouver, BC, Canada.

Chen F, Manning KW, Lemone MA, Trier SB, Alfieri JG, Roberts R, Tewari M, Niyogi D, Horst TW, Oncley SP, Basara JB, Blanken PD (2007) Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. J. Appl. Meteorol. Climatol. 46: 694–713. doi: 10.1175/JAM2463.1

Chen F, Kusaka H, Bornstein R, Ching J, Grimmond CSB, Grossman-Clarke S, Loridan T, Manning KW, Martilli A, Miao S, Sailor D, Salamanca FP, Taha H, Tewari M, Wang X, Wyszogrodzki AA, Zhang C (2011) The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. International Journal of Climatology, 31, 273-288, doi:10.1002/joc.2158

Ching, JKS (2013) A perspective on urban canopy layer modelling for weather, climate and air quality applications, Urban Climate, 3, 13-39.

Ching, J, Brown M, Burian S, Chen F, Cionco R, Hanna A, Hultgren T, McPherson T, Sailor D, Taha H, Williams D (2009) National urban database and access portal tool, Bull. American Meteorol. Soc., 90, 1157-1168, doi:10.1175/2009BAMS2675.1.

Chow WTL, Pope RL, Martin CA, Brazel AJ (2011) Observing and modeling the nocturnal park cool island of an arid city: horizontal and vertical impacts. Theoretical and applied climatology 103, 197-211.

Christen, A, Voogt R (2004) Energy and radiation balance of a central European city, Int. J.

Climatol., 24, 1395-1421, doi:10.1002/joc.1074.

Cleugh HA, Oke TR (1986) Suburban-rural energy balance comparisons in summer for Vancouver, B.C. Boundary- Layer Meteorology 36, 351-369.

Coceal O, Belcher SE (2004) A canopy model of mean winds through urban areas. Q. J. R. Meteorol. Soc. 130, 1349-1372.

Coulthard TJ, Frostick LE (2010) The Hull floods of 2007: implications for the governance and management of urban drainage systems. Journal of flood risk management 3, 223-231.

Coutts, AM, Beringer J, Tapper NJ (2007a) Characteristics influencing the variability of urban CO2 fluxes in Melbourne, Australia, Atmos. Environ., 41, 51-62.

Coutts, AM, Beringer J, Tapper NJ (2007b) Impact of increasing urban density on local climate: spatial and temporal variations in the surface energy balance in Melbourne, Australia, J. Appl. Meteorol., 47, 477-493.

Cox PM, Betts RA, Jones CD, Spall SA, Totterdell IJ (2000) Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature 408, 184-187.

Crawford B, Grimmond CSB, Christen A (2011) Five years of carbon dioxide fluxes measurements in a highly vegetated suburban area. Atmos. Environ. 45: 895-905.

Cullen M J P (1993) The unified forecast/climate model, Meteorol. Mag., 122, 81–94.

Dandou A., Tombrou M, Akylas E, Soulakellis N, Bossioli E (2005) Development and evaluation of an urban parameterization scheme in the Penn State/NCAR Mesoscale model (MM5). J Geophys Res, 110: D10102. doi:10.1029/2004JD005192.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljarrs ACM, van de Berg L, Bidlot J, Bormann N., Delsol C., Dragani R., Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C., de Rosnay P, Tavolato C, Thépaut J-N, Vitart F (2011) The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597

de Goncalves LGG, Shuttleworth WJ, Burke EJ, Houser P, Tall DL, Rodell M, Arsenault K (2006) Toward a sourth America land data assimilation system: Aspects of land surface model spin-up using the simplified simple biosphere. J. Geo. Res. – Atmospheres, 111, D17110, doi:10.1029/2005JD006297

Dirmeyer PA, Gao X, Zhao M, Zhichong G, Oki T, Hanasaki N (2006) GSWP-2 multimodel analysis and implications for our perception of the land surface. Bull American Meteorol Soc 87: 1381-1397.

Di Sabatino S, Solazzo E, Paradisi P, Britter,R (2008) A Simple Model for Spatially-averaged Wind Profiles Within and Above an Urban Canopy. Boundary-Layer Meteorology 127, 131-151

Dupont S, Mestayer PG (2006) Parameterisation of the urban energy budget with the submesoscale soil model. J Appl Meteorol Climatol, 45: 1744–1765.

Dupont S, Mestayer PG, Guilloteau E, Berthier E, Andrieu H (2006) Parameterisation of the urban water budget with the submesoscale soil model. J Appl Meteorol Climatol, 45: 624–648.

Ellefsen R (1985) Urban terrain zone characteristics. Tech. Monogr., No. 18-87, U.S. Army Engineering Laboratory, Aberdeen Proving Ground, MD, 350 pp.

Ellefsen, R (1991), Mapping and measuring buildings in the canopy boundary layer in ten U.S. cities. Energy and Buildings, 16, 1025-1049.

Essery RLH, Best MJ, Betts RA, Cox PM, Taylor CM (2003) Explicit representation of subgrid heterogeneity in a GCM land surface scheme. J Hydrometeorol, 4: 530–543.

Essery R, Rutter N, Pomeroy J, Baxter R, Stähli M, Gustafsson D, Barr A, Bartlett P, Elder K (2009) SNOWMIP2: An Evaluation of Forest Snow Process Simulations. Bulletin of the American Meteorological Society 90, 1120–1135.

Faroux S, Kaptue Tchuente AT, Roujean J-L, Masson V, Martin E, Le Moigne P (2013) ECOCLIMAP-II/Europe: a twofold database of ecosystems and surface parameters at 1 km resolution based on satellite information for use in land surface, meteorological and climate models, Geosci. Model Dev., 6, 563-582, doi:10.5194/gmd-6-563-2013.

Fortuniak K (2003) A slab surface energy balance model (SUEB) and its application to the study on the role of roughness length in forming an urban heat island. Acta Universitatis Wratislaviensis, 2542, 368-377.

Fortuniak K, Offerle B, Grimmond CSB (2004) Slab surface energy balance scheme and its application to parameterisation of the energy fluxes on urban areas. NATO ASI, Kiev, Ukraine; 82–83. Available from: www.met.rdg.ac.uk/urb met/NATO ASI/talks.html (Last accessed 4–15 May 2010).

Fortuniak K, Offerle B, Grimmond CSB (2005) Application of a slab surface energy balance model to determine surface parameters for urban areas. Lund Electronic Reports in Physical Geography, 5: 90–91.

Francis RA, Chadwick MA (2013) Urban Ecosystems: Understanding the Human Environment. Routledge, 220pp.

Garratt JR (1992) The atmospheric boundary layer. Cambridge University Press.

Goodwin NR, Coops NC, Tooke TR, Christen A, Voogt JA (2009) Characterizing urban surface cover and structure with airborne lidar technology. Can. J. Remote Sens., 35, 297-309.

Grimmond CSB (1992) The suburban energy balance: Methodological considerations and results for a mid-latitude west coast city under winter and spring conditions. International Journal of Climatology 12, 481-497.

Grimmond CSB (1996) Dynamically determined parameters for urban energy and water

exchange modelling. GIS and Environmental Modelling: Progress and Research Issues, M.F. Goodchild et al., Eds., GIS World Books, 305-309.

Grimmond CSB, Oke TR (1986) Urban water balance. 2. Results from a suburb of Vancouver, British Columbia. Water Res. Research 22: 1404-1412.

Grimmond CSB, Oke TR (1991) An evaporation-interception model for urban areas. Water Res. Research 27, 1739-1755.

Grimmond, CSB, Souch C (1994) Surface description for urban climate studies: a GIS based methodology. Geocarto. International, 9, 47-59.

Grimmond CSB, Oke TR (1995) Comparison of heat fluxes from summertime observations in the suburbs of four North American cities. J Appl Meteorol 34: 873-889.

Grimmond CSB, Oke TR (1999a) Heat storage in urban areas: Local-scale observations and evaluation of a simple model. Journal of Applied Meteorology 38: 922-940

Grimmond CSB, Oke TR (1999b) Aerodynamic properties of urban areas derived from analysis of surface form. Journal of Applied Meteorology 38: 1262-1292

Grimmond CSB, Oke TR (2002) Turbulent heat fluxes in urban areas: observations and local-scale urban meteorological parameterization scheme (LUMPS). J Appl Meteorol, 41: 792–810.

Grimmond CSB, Oke TR, Cleugh HA (1993) The role of 'rural' in comparisons of observed suburban-rural flux differences. Int. Assoc. Hydrolog. Sci. Publ. 212: 165-174.

Grimmond CSB, Souch C, Hubble MD (1996) Influence of tree cover on summertime surface energy balance fluxes, San Gabriel Valley, Los Angeles. Clim Res 6: 45-57.

Grimmond, CSB, Salmond JA, Oke TR, Offerle B, Lemonsu A (2004a) Flux and turbulence measurements at a densely built-up site in Marseille: Heat, mass (water and carbon dioxide), and momentum. J. Geophys. Res. Atmos., 109, D24101, doi:10.1029/2004jd004936

Grimmond CSB, Su H-B, Offerle B, Crawford B, Scott S, Zhong S, Clements C (2004b) Variability of sensible heat fluxes in a suburban area of Oklahoma City. Paper J7.2 in Symp. on Planning, nowcasting and forecasting in the urban zone, Eighth Symp. on Integrated observing and assimilation systems for atmosphere, oceans, and land surface. Am.Meteorol. Soc: Boston, MA.

Grimmond, CSB, Blackett M, Best MJ, Barlow J, Baik J-J, Belcher SE, Bohnenstengel SI, Calmet I, Chen F, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Shashua-Bar L, Steeneveld G-J, Trombou M, Voogt MJ, Young D, Zhang N (2010) The international urban energy balance models comparison project: first results from phase 1, J. Appl. Meteorol. Climatol., 49, 1268-1292, doi: 10.1175/2010JAMC2354.1.

Grimmond, CSB, Blackett M, Best MJ, Baik J-J, Belcher SE, Beringer J, Bohnenstengel SI,

Calmet I, Chen F, Coutts A, Dandou A, Fortuniak K, Gouvea ML , Hamdi R, Hendry M, Kanda M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Ooka R, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Steeneveld G-J, Trombou M, Voogt J, Young D, Zhang N (2011) Initial results from phase 2 of the international urban energy balance model comparison, Int. J. Climatol., 30, 244-272, doi:10.1002/joc.2227.

Hallenbeck M, Rice M, Smith B, Cornell-Martinez C, Wilkinson J (1997) Vehicle volume distribution by classification. Washington State Transportation Center, University of Washington, 54 pp. [Available from Washington State Transportation Center, University of Washington, 1107 NE 45th St. Suite 535, Seattle WA 98105]

Hamdi R, Schayes G (2007) Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities. Atmos Chem Phys 7: 4513–4526.

Hamdi, R, Masson V (2008) Inclusion of a drag approach in the Town Energy Balance (TEB) scheme: offline 1-D evaluation in a street canyon, J. Appl. Meteorol. Climatol., 47, 2627-2644.

Hamdi R, Degrauwe D, Duerinckx A, Cedilnik J, Costa V, Dalkilic T, Essaouini K, Jerczynki M, Kocaman F, Kullmann L, Mahfouf J-F, Meier F, Sassi M, Schneider S, Váňa F, Termonia P (2014) Evaluating the performance of SURFEXv5 as a new land surface scheme for the ALADINcy36 and ALARO-0 models. Geosci. Model Dev. 7: 23-39. doi: 10.5194/gmd-7-23-2014.

Hamilton IG, Davies M, Steadman P, Stone A, Ridley I, Evans S (2009) The significance of the anthropogenic heat emissions of London's buildings: A comparison against captured shortwave solar radiation. Building and Environ. 44, 807-817.

Harman IN, Belcher SE (2006) The surface energy balance and boundary layer over urban street canyons. Q J R Meteorol Soc, 132: 2749–2768.

Harman IN, Barlow JF, Belcher SE (2004a) Scalar fluxes from urban street canyons. Part II: model. Boundary-Layer Meteorol, 113: 387–410.

Harman IN, Best MJ, Belcher SE (2004b) Radiative exchange in an urban street canyon. Boundary-Layer Meteorol, 110: 301–316.

Harris I, Jones PD, Osborn TJ, Lister DH (2013) Updated high-resolution grids of monthly climatic observations - The CRU TS3.10 dataset. Int. J. Climatol., 34, 623–642, doi:10.1002/joc.3711.

Henderson-Sellers A (1996) Soil moisture simulation: Achievements of the RICE and PILPS intercomparison workshop and future direction. Global and planetary change 13, 99-115.

Henderson-Sellers A, Yang ZL, Dickenson RE (1993) The project for intercomparison of land-surface parameterization schemes. Bulletin of the American Meteorological Society 74: 1335–1349.

Henderson-Sellers A, Irannejad P, McGuffie K, Pitman A (2003) Predicting land-surface climates-better skill or moving targets?. Geophysical Research Letters 30: 1777, DOI:10.1029/2003GL017387.

Herbst, J, Mason K, Byard RW, Bilbert JD, Charlwood C, Heath KJ, Winskog C, Langlois NEI (2014) Heat-related deaths in Adelaide, South Australia: Review of the literature and case findings – An Australian perspective, J. Forensic and Legal Medicine, 22, 73-78. doi:10.1016/j.jflm.2013.12.018

Hollinger DY, Richardson AD (2005) Uncertainty in eddy covariance measurements and its application to physiological models, Tree Physiol., 25, 873-885.

Howard L (1833) Climate of London deduced from meteorological observations, made in the metropolis, and at various places around it. Vol. 1-3. London: Harvey and Darton.

IPCC (2014) Working Group I Contribution to the IPCC Fifth Assessment Report (AR5), Climate Change 2013: The Physical Science Basis. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Iamarino M, Beevers S, Grimmond CSB (2012) High-resolution (space, time) anthropogenic heat emissions: London 1970-2025. Int J Climatol 32: 1754-1767. doi:10.1002/joc.2390

Ichinose T, Shimodozono K, Hanaki K (1999) Impact of anthropogenic heat on urban climate in Tokyo. Atmos Environ 33: 3897-3909

Irranejad P, Henderson-Sellers A, Sharmeen S (2003) Importance of land-surface parameterization for latent heat simulation in global atmospheric models. Geophysical Research Letters 30: 1904, DOI:10.1029/2003/GL018044.

Jackson TL, Feddema JJ, Oleson KW, Bonan GB, Bauer JT (2010) Parameterization of urban characteristics for global climate modelling. Annals of the Association of American Geographers 100, 848-865.

Järvi L, Grimmond CSB, Christen A (2011) The surface urban energy and water balance scheme (SUEWS): Evaluation in Los Angeles and Vancouver. J Hydrol 411: 219-237. doi:10.1016/j.jhydrol.2011.10.001

Järvi L, Grimmond CSB, Taka M, Nordbo A., Setälä H, Strachan IB (2014) Development of the surface urban energy and water balance scheme (SUEWS) for cold climate cities. Geosci. Model Dev. 7: 1691-1711. doi:10.5194/gmd-7-1691-2014.

Kanda M, Kawai T, Kanega M, Moriwaki R, Narita K, Hagishima A (2005a) A simple energy balance model for regular building arrays. Boundary-Layer Meteorol, 116: 423–443.

Kanda M, Kawai T, Nakagawa K (2005b) A simple theoretical radiation scheme for regular building arrays. Boundary-Layer Meteorol, 114: 71–90.

Kawai T, Kanda M, Narita K, Hagishima A (2007) Validation of a numerical model for urban energy-exchange using outdoor scalemodel measurements. Int J Climatol, 27: 1931–1942.

Kawai T, Ridwan MK, Kanda M (2009) Evaluation of the simple urban energy balance model

using 1-yr flux observations at two cities. J Appl Meteorol Climatol, 48: 693–715.

Kawamoto Y, Ooka R (2006) Analysis of the radiation field at pedestrian level using a meso-scale meteorological model incorporating the urban canopy model. In ICUC-6, Göteborg, Sweden, 12–16 June 2006.

Kawamoto Y, Ooka R (2009a) Accuracy validation of urban climate analysis model using MM5 incorporating a multi-layer urban canopy model. In ICUC-7, Yokohama, Japan, 28 June–3 July 2009.

Kawamoto Y, Ooka R (2009b) Development of urban climate analysis model using MM5 Part 2 – incorporating an urban canopy model to represent the effect of buildings. J Environ Eng (Transactions of AIJ) 74(642): 1009–1018 (in Japanese).

King T, Grimmond CSB (1997) Transfer mechanisms over an urban surface for water vapor, sensible heat, and momentum. In Preprints, 12th Conf. on Boundary layers and turbulence, Vancouver, BC, Canada. Am. Meteorol. Soc: Boston, M.A. pp 455-456

Kondo H, Liu FH (1998) A study on the urban thermal environment obtained through a one-dimensional urban canopy model, J Jpn Soc Atmos Environ. 33, 179-192 (in Japanese)

Kondo H, Genchi Y, Kikegawa Y, Ohashi Y, Yoshikado H, Komiyama H (2005) Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: structure of the urban canopy model and its basic performance. Boundary-Layer Meteorol, 116: 395–421.

Koster RD, Suarez MJ (1992) Modelling the land surface boundary in climate models as a composite of independent vegetation stands. J Geophys Res (Atmos) 97: 2697-2715.

Koster RD, Sud YC, Guo Z, Dirmeyer PA, Bonan G, Oleson KW, Chan E, Verseghy D, Cox P, Davies H, Kowalczyk E, Gordon CT, Kanae S, Lawrence D, Piu P, Mocko D, Lu C-H, Mitchell K, Malyshev S, McAvaney B, Oki T, Yamada T, Pitman A, Taylor CM, Vasic R, Xue Y (2006) GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. Journal of Hydrometeorology 7, 611–625.

Kotthaus S, Grimmond CSB (2012) Identification of micro-scale anthropogenic CO_2, heat and moisture sources – Processing eddy covariance fluxes for a dense urban environment. Atmos. Environ. 57, 301-316. doi:10.1016/j.atmosenv.2012.04.024.

Kotthaus S, Grimmond CSB (2014a) Energy exchange in a dense urban environment – Part I: Temporal variability of long-term observations in central London. Urban Climate, 10, 261-280, doi:10.1016/j.uclim.2013.10.002

Kotthaus S., Grimmond CSB (2014b) Energy exchange in a dense urban environment - Part II: Impact of spatial heterogeneity of the surface, Urban Climate, 10, 281-307, hppt://dx.doi/org/10.1016/j.uclim.2013.10.001.

Krayenhoff ES, Voogt JA (2007) A microscale three-dimensional urban energy balance model for studying surface temperatures, Boundary-Layer Meteorol., 123, 433-461.

Kusaka H, Kondo H, Kikegawa Y, Kimura F (2001) A simple singlelayer urban canopy model for atmospheric models: comparison with multi-layer and slab models, Boundary-Layer Meteorol., 101, 329-358.

Laaidi K, Zeghnoun A, Dousset B, Bretin P, Vandentorren S, Giraudet E, Beaudeau P (2012) The impact of heat islands on mortality in Paris during the August 2003 heat wave, Environ. Health Perspectives, 120, 254-259, doi:10.1289/ehp.1103532.

Landsberg HE (1981) The urban climate. International Geophysics Series 28, pp. 275. New York: Academic Press.

Lee S-H, Park S-U (2008) A vegetated urban canopy model for meteorological and environmental modelling, Boundary-Layer Meteorol., 126, 73-102.

Lee SH , Baik JJ (2010) Statistical and dynamical characteristics of the urban heat island intensity in Seoul. Theoretical and Applied Climatology 100, 227-237.

Lemonsu A, Grimmond CSB, Masson V (2004) Modelling the surface energy balance of an old Mediterranean city core. J Appl Meteorol, 43: 312–327.

Lemonsu A, Belair S, Mailhot J (2009) The new Canadian urban modelling system: Evaluation for two cases from the joint urban 2003 Oklahoma City experiment. Boundary- Layer Meteorol 113: 47-70. doi: 10.1007/s10546-009-9414-2.

Leuning R, van Gorsel E, Massman WJ, Isacc PR (2012) Reflections on the surface energy imbalance problem. Agricultural and Forest Meteorol. 156, 65-74. doi:10.1016/j.agrformet.2011.12.002.

Li CB, Zhou JJ, Cao YJ, Zhong J, Liu Y, Kang CQ, Tan Y (2014) Interaction between urban microclimate and electric air-conditioning energy consumption during high temperature season, Applied Energy, 117, 149-156, doi:10.1016/j.apenergy.2013.11.057.

Lindberg F, Grimmond CSB (2010) Continuous sky view factor mpas from high resolution urban digital elevation models. Climate Research 42, 177-183.

Lindberg F, Grimmond CSB (2011) Nature of vegetation and building morphology characteristics across a city: Influence on shadow patterns and mean radiant temperatures in London. Urban Ecosyst., 14, 617-623. doi:10.1007/s11252-011-0184-5.

Lindberg F, Grimmond CSB, Yogeswaran N, Kotthaus S, Alen L (2013) Impacts of city changes and weather on anthropogenic heat flux in Europe 1995-2015, Urban Climate, 4, 1-15. http://dx.doi.org/10.1016/j.uclim.2013.03.002.

Linden J (2011) Nocturnal cool island in the Sahelian city of Ouagadougou, Burkina Faso. International Journal of Climatology 31, 605-620.

Loridan T, Grimmond CSB (2012a) Characterization of energy flux partitioning in urban environments: links with surface seasonal properties. J. Appl. Meteorol. Climatol. 51, 219-241 doi: 10.1175/JAMC-D-11-038.1

Loridan T, Grimmond CSB (2012b) Multi-site evaluation of an urban land-surface model: intra-

urban heterogeneity, seasonality, and parameter complexity requirements. Q J R Meteorol Soc, 138: 1094–1113. doi:10.1002/qj.963

Loridan T, Grimmond CSB, Grossman-Clarke S, Chen F, Tewari M, Manning K, Martilli A, Kusaka H, Best M (2010) Trade-offs and responsiveness of the single-layer urban canopy parameterization in WRF: an offline evaluation using the MOSCEM optimization algorithm and field observations. Q J R Meteorol Soc, 136: 997–1019. doi:10.1002/qj.614.

Loridan T, Grimmond CSB, Offerle BD, Young DT, Smith T, Järvi L, Lindberg F (2011) Local-scale urban meteorological parameterization scheme (LUMPS): Longwave radiation parameterization and seasonality-related developments. J Appl Meteorol Climatol, 50: 185-202. doi:10.1175/2010JAMC2474.1

Luhar AK, Thatcher M, Hurley PJ (2014) Evaluating a building-averaged urban surface scheme in an operational mesoscale model for flow and dispersion, Atmos. Environ., 88, 47-58, doi:10.1016/j.atmosenv.2014.01.059.

Macdonald RW (2000) Modelling the mean velocity profile in the urban canopy layer. Boundary-Layer Meteorology 97, 25-45

Manabe S (1969) Climate and the ocean circulation: 1, The atmospheric circulation and the hydrology of the earths' surface. Mon Wea Rev 97: 739-805

Martilli A, Clappier A, Rotach MW (2002) An urban surface exchange parameterisation for mesoscale models. Boundary-Layer Meteorol, 104: 261–304.

Masson V (2000) A physically-based scheme for the urban energy budget in atmospheric models. Boundary-Layer Meteorol, 41: 1011–1026.

Masson V, Grimmond CSB, Oke TR (2002) Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities. J Appl Meteorol, 41: 1011–1026.

Masson V, Champeaux J-L, Chauvin F, Meriguet C, Lacaze R (2003) A global database of land surface parameters at 1-km resolution in meteorological and climate models. J. Climate 16(9), 1261-1282

McCarthy MP, Best MJ, Betts RA (2010) Climate change in cities due to global warming and urban effects, Geophys. Res. Letters, 37, L09705, doi:10.1029/2010GL042845.

Meier F, Scherer D (2012) Spatial and temporal variability of urban tree canopy temperature during summer 2010 in Berlin, Germany. Theoretical and Applied Climatol., 110, 373-384. doi: 10.1007/s00704-012-0631-0.

Mestayer PG, Durand P, Augustin P, Bastin S, Bonnefond J-M, Béenech B, Campistron B, Coppalle A, Delbarre H, Dousset B, Drobinski P, Druilhet A, Fréjafon E, Grimmond CSB, Groleau D, Irvine M, Kergomard C, Kermadi S, Lagouarde J-P, Lenonsu A., Lohou F, Long N, Masson V, Moppert C, Noilhan J, Offerle B, Oke TR, Pigeon G, Puygrenier V, Roberts S, Rosanti J-M, Saïd F, Salmond J, Talbaut M, Voogt J (2005) The urban boundary-layer field

campaign in Marseille (UBL/CLU-ESCOMPTE) : Set-up and first results. Boundary-Layer Meteorol 114, 315-365.

Myrup LO (1969) A numerical study of the urban heat island. J. Appl. Meteorol. 8, 896-907.

Neal JC, Bates PD, Fewtrell TJ, Hunter NM, Wilson MD, Horritt MS (2009) Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. Journal of Hydrology 368, 42-55.

New M, Hulme M, Jones P (1999) Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. J. Climate, 12, 829–856.

New M, Hulme M, Jones P (2000) Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. J. Climate, 13, 2217–2238.

Newton T (1999) Energy balance fluxes in a subtropical city: Miami, FL. M.S. thesis, Dept. of Geography, University of British Columbia, Vancouver, BC, Canada, 140 pp.

Newton T, Oke TR, Grimmond CSB, Roth M (2007) The suburban energy balance in Miami, Florida. Geografiska Annaler Series A – Phys Geogr 89A: 331-347.

Notaro M, Liu Z, Gallimore RG, Williams JW, Gutzler DS, Collins S (2010) Complex seasonal cycle of ecohydrology in the Southwest United States. J Geophys Res 115: G04034. doi:10.1029/2010JG001382.

Nunez M, Oke TR (1977) The energy balance of an urban canyon. J Appl Meteorol, 16: 11-19.

Offerle B., C.S.B. Grimmond, T.R. Oke (2003), Parameterization of net all-wave radiation for urban areas. J Appl Meteorol, 42: 1157–1173.

Offerle B, Grimmond CSB, Fortuniak K (2005a) Heat storage and anthropogenic heat flux in relation to the energy balance of a central European city centre. Int J Climatol 25: 1405-1419. doi:10.1002/joc.1198

Offerle B, Jonsson P, Eliasson I, Grimmond CSB (2005b) Urban modification of the surface energy balance in the West African Sahel: Ouagadougou, Burkina Faso. J. Climatol. 18: 3983-3995.

Offerle B, Grimmond CSB, Fortuniak K, Kłysik K, Oke TR. (2006a) Temporal variations in heat fluxes over a central European city centre. Theorectical and Applied Climatol. 84: 103-115.

Offerle B, Grimmond CSB, Fortuniak K, Pawlak W (2006b) Intraurban differences of surface energy fluxes in a central European city. J. Appl. Meteorol. Climatol. 45: 125-136.

Oke TR (1987) Boundary layer climates. Routeldge, 2nd edition.

Oke TR (1988) The urban energy balance. Progress in Physical Geography, 12, 471-508.

Oke TR (2004) Initial guidance to obtain representative meteorological observations at urban sites. Instruments and Observing Methods Report 81, WMO/TD 1250, 51 pp.

Oke TR (2006) Towards better scientific communication in urban climate. Theor. Appl. Climatol. 84: 179-190. doi: 10.1007/s00704-005-0153-0

Oke TR, Hay JE (1998) The climate of Vancouver. 2nd Edition, BC Geographical Series, No. 50, Vancouver, 84 pp, ISBN 919478 74 3.

Oke TR, Spronken-Smith A, Jauregui E, Grimmond CSB (1999) The energy balance of central Mexico City during the dry season. Atmos Environ 33: 3919-3930.

Oleson,KW, Bonan GB, Feddema J, Vertenstein M, Grimmond CSB (2008a) An urban parameterization for a global climate model: 1. Formulation and evaluation for two cities, J. Appl. Meteorol. Climatol., 47, 1038-1060.

Oleson KW, Bonan GB, Feddema J, Vertenstein M (2008b) An urban parameterization for a global climate model: 2. Sensitivity to input parameters and the simulated heat island in offline simulations. J Appl Meteorol Climatol, 47: 1061–1076.

Oleson, KW, Bonan GB, Feddema J, Jackson T (2011) An examination of urban heat island characteristics in a global climate model, Int. J. Climatol., 31, 1848-1865. doi:10.1002/joc.2201.

Pawlak W, Fortuniak K, Siedlecki M (2011) Carbon dioxide flux in the centre of Łódź, Poland – analysis of a 2-year eddy covariance measurement data set. Int. J. Climatol. 31: 232-243. doi:10.1002/joc.2247

Pickett STA., Cadenasso ML, Grove JM, Groffman PM, Band LE, Boone CG, Burch Jr. WR, Grimmond CSB, Hom J, Jenkins JC, Law NL, Nilon CH, Pouyat RV, Szlavecz K, Warren PS, Wilson MA (2008) Beyond urban legends: an emerging framework of urban ecology, as illustrated by the Baltimore Ecosystem Study. BioScience. 58(2):139-150

Pigeon G, Lemonsu A, Grimmond CSB, Durand P, Thouron O, Masson V (2007a) Divergence of turbulent fluxes in the surface layer: case of a coastal city. Boundary-Layer Meteorology 124, 269-290.Roth M. 2000. Review of atmospheric turbulence over cities. Q J R Meteorol Soc 126, 941-990

Pigeon G, Legain D, Durand P, Masson V (2007b) Anthropogenic heat release in an old European agglomeration (Toulouse, France). Int J Climatol 27: 1969-1981

Pigeon G, Moscicki MA, Voogt JA, Masson V (2008) Simulation of fall and winter surface energy balance over a dense urban area using the TEB scheme. Meteorol Atmos Phys, 102: 159–171.

Polcher J, Laval K, Dümenil L, Lean J and Rowntree P (1996). Comparing three land surface schemes used in general circulation models. Journal of Hydrology 180, 373-394. doi:10.1016/0022-1694(95)02886-2

Porson A, Harman IN, Bohnenstengel SI, Belcher SE (2009) How many facets are needed to represent the surface energy balance of an urban area? Boundary-Layer Meteorol 132: 107–128.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE (2010) Implementation of a new urban energy budget scheme in the MetUM. Part II. Validation against observations and model intercomparison. Q J R Meteorol Soc, 136: 1530-1542.

POST (2008) Renewable energy in a changing climate. Parliamentary Office of Science and Technology, Postnote 315, October 2008. http://www.parliament.uk/documents/post/postpn315.pdf (last accessed 12/05/2011)

Prata AJ (1996) A new long-wave formula for estimating downward clear-sky radiation at the surface. Q J R Meteorol. Soc. 112, 1127-1151.

Quah AKL, Roth M (2011) Diurnal and weekly variation of anthropogenic heat emissions in a tropical city, Singapore. Atmos Environ 46: 92-103

Radhi, H, Sharples S (2013) Quantifying the domestic electricity consumption for air-conditioning due to urban heat islands in hot arid regions, Applied Energy, 112, 371-380, doi:10.1016/j.apenergy.2013.06.013.

Ross SL, Oke TR (1988) Tests of three urban energy balance models. Boundary-Layer Meteorol 44: 73-96.

Ryu Y-H, Baik J-J, Lee S-H (2011) A new single-layer urban canopy model for use in mesoscale atmospheric models. J Appl Meteorol Climatol, 50: 1773-1794. doi: 10.1175/2011JAMC2665.1

Ryu, Y-H, Baik J-J (2013) Effects of anthropogenic heat on ozone air quality in a megacity. Atmos. Environ., 80, 20-30, doi:10.1016/j.atmosenv.2013.07.053.

Saaroni H, Ben-Dor E, Bitan A, Potchter O (2000) Spatial distribution and microscale characteristics of the urban heat island in Tel_Aviv, Israel. Landscape and Urban Planning 48, 1-18.

Saha, MV, Davis RE, Hondula DM (2014) Mortality displacement as a function of heat event strength in 7 US cities, American J. Epidemiology, 179, 467-474, doi:10.1093/aje/kwt264.

Sailor DJ (2011) A review of methods for estimating anthropogenic heat and moisture emissions in the urban environment. Int J Climatol 31: 189-199.

Sailor, DJ, Lu L (2004) A top-down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. Atmos. Environ., 38, 2737-2748, doi:10.1016/j.atmosenv.2004.01.034.

Salamanca F, Martilli A (2010) A new Building Energy Model coupled with an Urban Canopy Parameterization for urban climate simulations – part II. Validation with one dimension off-line simulations. Theor Appl Climatol, 99: 345–356.

Salamanca F, Krayenhoff ES, Martilli A (2009) On the derivation of material thermal properties representative of heterogeneous urban neighbourhoods. J Appl Meteorol Climatol, 48: 1725–1732.

Salamanca F, Krpo A, Martilli A, Clappier A (2010) A new building energy model coupled

with an urban canopy parameterization for urban climate simulations – part I. Formulation, verification, and sensitivity analysis of the model. Theor Appl Climatol, 99: 345-356. doi: 10.1007/s00704-009-0142-9.

Sarrat C, Lemonsu A, Masson V, Guedalla D (2006) Impact of urban heat island on regional atmospheric pollution, Atmos. Environ., 40, 1743-1758, doi:10.1016/j.atmosenv.2005.11.037.

Schlosser CA, Slater AG, Robock A, Pitman AJ, Vinnikov KY, Henderson-Sellers A, Speranskaya NA, Mitchell K (2000) Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(D). Monthly weather review 128, 301-321.

Schmid HP (1994) Source areas for scalars and scalar fluxes. Boundary Layer Meteorol. 67, 293-318.

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Ziese M, Rudolf B (2013) GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. Theor. Appl. Climatol. 115: 15-40. doi: 10.1007/s00704-013-0860-x

Sheng J and Zwiers F (1998) An improved scheme for time-dependent boundary conditions in atmospheric general circulation models, Climate Dynamics, 14, 609-613.

Shrestha R, Houser P (2010) A heterogeneous land surface model initialization study. J. Geo. Res. – Atmospheres, 115, D19111, doi:10.1029/2009JD013252

Soux A, Voogt JA, Oke TR (2004) A model to calculate what a remote sensor 'sees' of an urban surface. Boundary-Layer Meteorol. 111, 109-132.

Spronken-Smith RA, Oke TR (1998) The thermal regime of urban parks in two cities with different summer climates. International Journal of remote Sensing 19, 2085-2104.

Stewart, ID, Oke TR (2012) Local climate zones for urban temperature studies, Bull. American Meteorol. Soc., 93, 1879-1900, doi:10.1175/BAMS-D-11-00019.1.

Steyn DG (1985) An objective model to achieve closure of over determined surface energy budgets. Boundary-Layer Meteorol. 33, 303-311.

Svensson MK (2004) Sky view factor analysis – implications for urban temperature differences. Meteorological Applications 11, 201-211.

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106: 7183-7192.

Taylor CM, Parker DJ, Kalthoff N, Gaertner MA, Philippon N, Bastin S, Harris PP, Boone A, Guichard F, Agusti-Panareda A, Baldi M, Cerlini P, Descroix L, Douville H, Flamant C, Grandpeix JY, Polcher J (2011) New perspectives on land-atmosphere feedbacks from the African Monsoon Multidisciplinary Analysis. Atmospheric science letter 12, 38-44.

Vesala T, Järvi L, Launiainen S, Sogachev A, Rannik Ü, Mammarella I, Siivola E, Keronen P, Rinne J, Riikonen A, Nikinmaa E (2008) Surface-atmosphere interactions over complex

urban terrain in Helsinki, Finland. Tellus 60B: 188-199, doi:10.1111/j.1600-0889.2007.00312.x.

Viterbo P, Beljaars ACM (1995) An improved land-surface parameterization scheme in the ECMWF model and its validation. Journal of Climate 8, 2716-2748.

Voogt JA, Grimmond CSB (2000) Modeling surface sensible heat flux using surface radiative temperatures in a simple urban area. J Appl Meteorol. 39: 1679–1699.

Weedon GP, Gomes S, Viterbo P, Shuttleworth WJ, Blyth E, Österle H, Adam JC, Bellouin N, Boucher O, Best M (2011) Creation of the WATCH Forcing Data and its use to assess global and regional reference crop evaporation over land during the twentieth century. J. Hydrometeorol. 12: 823-848. doi:10.1175/2011JHM1369.1

Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P (2014) The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res. 50, doi:10.1002/2014WR015638

World Urbanization Prospects (2009) United Nations, Department of Economic and Social Affairs, Population Division. File 2: Percentage of Population Residing in Urban Areas by Major Area, Region and Country, 1950-2050. POP/DB/WUP/Rev.2009/1/F2, http://esa.un.org/unpd/wup/index.htm

Appendix A: The international urban energy balance models comparison project: First results from phase 1.

The International Urban Energy Balance Models Comparison Project: First Results from Phase 1

C. S. B. GRIMMOND,[a] M. BLACKETT,[a] M. J. BEST,[b] J. BARLOW,[c] J.-J. BAIK,[d] S. E. BELCHER,[c] S. I. BOHNENSTENGEL,[c] I. CALMET,[e] F. CHEN,[f] A. DANDOU,[g] K. FORTUNIAK,[h] M. L. GOUVEA,[a] R. HAMDI,[i] M. HENDRY,[b] T. KAWAI,[j] Y. KAWAMOTO,[k] H. KONDO,[l] E. S. KRAYENHOFF,[m] S.-H. LEE,[d] T. LORIDAN,[a] A. MARTILLI,[n] V. MASSON,[o] S. MIAO,[p] K. OLESON,[f] G. PIGEON,[o] A. PORSON,[b,c] Y.-H. RYU,[d] F. SALAMANCA,[n] L. SHASHUA-BAR,[q] G.-J. STEENEVELD,[r] M. TOMBROU,[g] J. VOOGT,[s] D. YOUNG,[a] AND N. ZHANG[t]

[a] *King's College London, London, United Kingdom*
[b] *Met Office, Exeter, United Kingdom*
[c] *University of Reading, Reading, United Kingdom*
[d] *Seoul National University, Seoul, South Korea*
[e] *Laboratoire de Mécanique des Fluides, CNRS-Ecole Centrale de Nantes, Nantes, France*
[f] *National Center for Atmospheric Research, Boulder, Colorado*
[g] *National and Kapodistrian University of Athens, Athens, Greece*
[h] *University of Łódź, Łódź, Poland*
[i] *Royal Meteorological Institute, Uccle, Belgium*
[j] *Ehime University, Matsuyama, Japan*
[k] *The University of Tokyo, Tokyo, Japan*
[l] *National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan*
[m] *University of British Columbia, Vancouver, British Columbia, Canada*
[n] *CIEMAT, Madrid, Spain*
[o] *CNRM-GAME Meteo France-CNRS, Toulouse, France*
[p] *IUM, CMA, Beijing, China*
[q] *Ben Gurion University of the Negev, Beer-Sheva, Israel*
[r] *Wageningen University, Wageningen, Netherlands*
[s] *University of Western Ontario, London, Ontario, Canada*
[t] *Nanjing University, Nanjing, China*

(Manuscript received 27 July 2009, in final form 4 February 2010)

ABSTRACT

A large number of urban surface energy balance models now exist with different assumptions about the important features of the surface and exchange processes that need to be incorporated. To date, no comparison of these models has been conducted; in contrast, models for natural surfaces have been compared extensively as part of the Project for Intercomparison of Land-surface Parameterization Schemes. Here, the methods and first results from an extensive international comparison of 33 models are presented. The aim of the comparison overall is to understand the complexity required to model energy and water exchanges in urban areas. The degree of complexity included in the models is outlined and impacts on model performance are discussed. During the comparison there have been significant developments in the models with resulting improvements in performance (root-mean-square error falling by up to two-thirds). Evaluation is based on a dataset containing net all-wave radiation, sensible heat, and latent heat flux observations for an industrial area in Vancouver, British Columbia, Canada. The aim of the comparison is twofold: to identify those modeling approaches that minimize the errors in the simulated fluxes of the urban energy balance and to determine the degree of model complexity required for accurate simulations. There is evidence that some classes of models perform better for individual fluxes but no model performs best or worst for all fluxes. In general, the simpler models perform as well as the more complex models based on all statistical measures. Generally the schemes have best overall capability to model net all-wave radiation and least capability to model latent heat flux.

Corresponding author address: Sue Grimmond, Environmental Monitoring and Modelling Group, Department of Geography, King's College London, London, WC2R 2LS, United Kingdom.
E-mail: sue.grimmond@kcl.ac.uk

DOI: 10.1175/2010JAMC2354.1

1. Introduction

The world's population has become increasingly urbanized: around 29% of the global population were urban dwellers in 1950, rising to 47% by 2000, and this proportion is predicted to rise to 69% by 2050 (UN 2009). Thus increasing numbers of people are impacted by weather and climate in urban areas. There is a growing requirement for accurate weather forecasts and climate change information within cities, and concurrent increases in computer capabilities allow greater spatial resolution within models. In combination, there is a greater proportion of the earth's surface being categorized as "urban" and there are a larger number of smaller grid boxes in atmospheric models in which urban areas need to be resolved.

The surface morphology (i.e., urban form) and presence of impervious building materials, sparseness of vegetation, and anthropogenic heat, water, and pollutant contributions each have a significant effect on the climate of urban regions, which lead to phenomena such as the urban heat island. Thus, effects of the urban surface on the fluxes of heat, moisture, and momentum need to be accounted for in the land-surface schemes used within numerical models, although the complexity of these schemes has to be balanced with their computational requirements. A fundamental aim of urban energy balance models is to accurately predict fluxes at the local scale (10^2–10^4 m). Some calculate additional terms including within-canyon air temperatures and wind speed, and facet surface temperature. A facet is a surface of the urban geometry that can be characterized by a single temperature and surface energy balance, and that can interact thermodynamically with other facets (e.g., a wall facet exchanging longwave radiation with the road facet; Fig. 1). The outputs from the model may be hourly or higher temporal resolution for the whole surface, or be facet/orientation-specific.

Models have been developed to incorporate urban features for different applications including global climate modeling (e.g., Oleson et al. 2008a,b); numerical weather prediction (e.g., Best 1998, 2005; Masson 2000; Chen et al. 2004; Harman and Belcher 2006; Liu et al. 2006); air quality forecasting (e.g., Martilli et al. 2003) and dispersion modeling (e.g., Hanna and Chang 1992, 1993); characterization of measurements (e.g., Krayenhoff and Voogt 2007); and water balance modeling (e.g., Grimmond et al. 1986; Grimmond and Oke 1991). Across these schemes a wide range of urban features are incorporated; the models have varying levels of complexity, and different fluxes modeled (Table 1; Figs. 1, 2).

In this paper, the methodology and initial results from the first international comparison of a broad range of urban land-surface schemes are presented. The requirements of a land-surface model from the perspective of an atmospheric model are considered; that is, surface fluxes of heat, moisture, and momentum. Thus, the fundamental requirement for the models to be included is that they simulate urban energy balance fluxes. The forcing data for the surface models are the same as that which would be provided by an atmospheric model; that is, the incoming shortwave ($K\downarrow$) and longwave fluxes ($L\downarrow$), air temperature, specific humidity, and the wind components. From these the outgoing radiative fluxes ($K\uparrow$, $L\uparrow$), net all-wave radiation (Q^*), turbulent sensible heat flux (Q_H), turbulent latent heat flux (Q_E), and net heat storage flux (ΔQ_S) are modeled. In this context, the net heat storage includes the energy storage within the buildings, the road and underlying soil, and for some models the air space within the street canyon (Grimmond and Oke 1999a). In the urban environment it is also useful to consider the anthropogenic heat flux (Q_F) in the surface energy balance (Oke 1988):

$$Q^* + Q_F = Q_H + Q_E + \Delta Q_S. \quad (1)$$

Features such as additional sources of energy (Q_F), presence of built and natural surfaces, the bluff body nature of the buildings, and existence of urban canyons, combine to change energy partitioning in urban areas. Thus significant modification to rural land parameterization schemes is needed. While many urban models have been evaluated against observational datasets (e.g., Grimmond and Oke 2002; Masson et al. 2002; Dupont and Mestayer 2006; Hamdi and Schayes 2007; Krayenhoff and Voogt 2007; Kawai et al. 2009), with some models even using the same observations, these comparisons have not been conducted in a controlled manner that allows robust model intercomparison. The objective here is to do just that: to undertake a staged and carefully controlled classification and comparison of urban energy balance models and their performance. An important objective also is to determine which approaches minimize the errors in the simulated fluxes.

2. The characteristics of urban energy balance models

Urban energy balance models can be classified in a number of ways (see also Grimmond et al. 2009a); for example, they vary in terms of the fluxes they calculate ("F" in Figs. 1 and 2). While all the models examined here calculate $K\uparrow$, $L\uparrow$, Q^*, and Q_H, some do not model either Q_E or the Q_F, and some model neither. Here, a series of features are used to classify the approaches

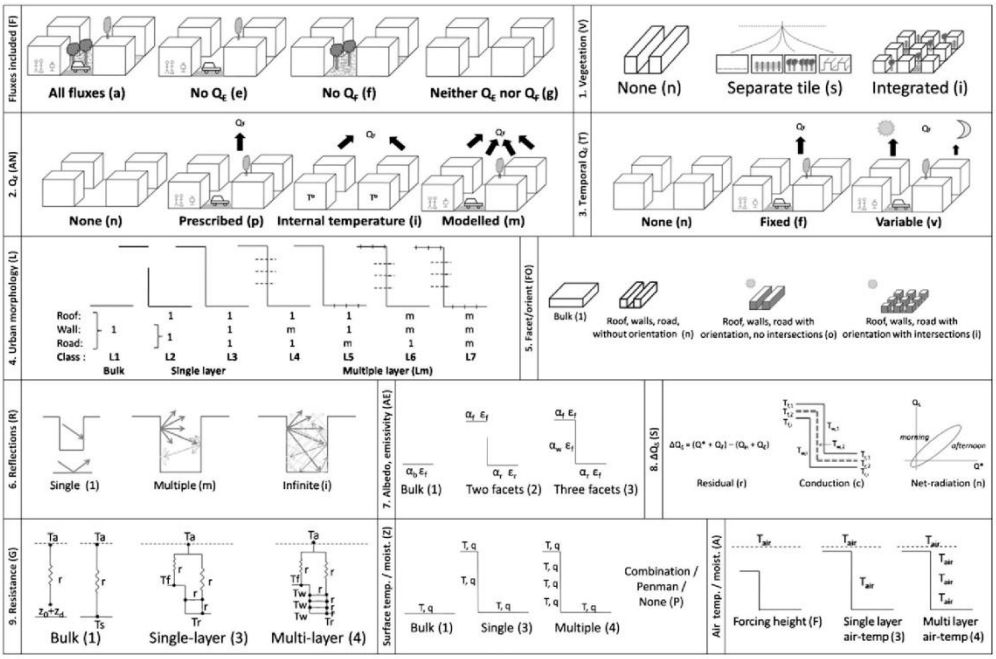

FIG. 1. Characteristics used to classify models (see Fig. 2).

taken. Figures 1 and 2 illustrate these, and the latter provides the numbers of models in each category. The illustrations also give each model class a reference in order to identify the category and its classification.

a. Vegetation and latent heat flux ("V" in Figs. 1 and 2)

A key decision in modeling an urban surface is whether or not vegetation (V) is simulated. A threefold classification is used here, where vegetation is

Vn: not considered,
Vs: modeled using a "tile" scheme to represent the surface heterogeneity (e.g., Essery et al. 2003) that does not interact with other surface types until the first atmospheric level of a mesoscale model (e.g., Best et al. 2006), and
Vi: "integrated" into the modeled urban surface.

The implication of not including vegetation is that there can be no latent heat except for periods immediately following rainfall. Some, even after rainfall, calculate no Q_E, whereas some account for dewfall and its later evaporation (Fig. 2b). For central business districts in many cities it may be reasonable to assume a negligible amount of vegetation and, hence, an absence of Q_E associated with vegetated surfaces. However, in residential areas (e.g., suburban North America) extensive fractions of the surface are vegetated, so the assumption of no urban Q_E is unrealistic. Moreover, in many locations, extensive street cleaning can result in water being available for evaporation despite the lack of vegetation (e.g., Mexico City, Mexico, Oke et al. 1999; Marseille, France, Grimmond et al. 2004).

The two classes of model that do incorporate vegetation differ in terms of the interactions which occur between the "built" and "vegetated" fractions (Figs. 2a–c). In the first case, "tiles" (Vs; Fig. 1), models typically take advantage of traditional land-surface schemes that have a wide variety of vegetation categories (e.g., Noilhan and Mahfouf 1996; Chen and Dudhia 2001; Essery and Clark 2003). Many have been extensively evaluated in the Project for Intercomparison of Land-Surface Parameterization Schemes (PILPS) (Henderson-Sellers et al. 1993, 2003; Irranejad et al. 2003) and other studies. Urban vegetation typically is more diverse than an individual vegetation class so a number of classes may be required (e.g., needleleaf and evergreen broadleaf trees)

TABLE 1. Urban energy balance models with the number of versions and number of groups utilizing each model participating in the comparison.

Code	Model Name	References	Versions	Groups
BEP02	Building Effect Parameterization	Martilli et al. (2002)	1	1
BEP_BEM08	BEP coupled with Building Energy Model	Martilli et al. (2002), Salamanca et al. (2009), Salamanca and Martilli (2009)	1	1
CLMU	Community Land Model–Urban	Oleson et al. (2008a,b)	1	1
GCTTC	Green Cluster Thermal Time Constant model	Shashua-Bar and Hoffman (2002, 2004)	1	1
IISUCM	Institute of Industrial Science Urban Canopy Model	Kawamoto and Ooka (2006, 2009a,b)	1	1
JULES	Joint U.K. Land Environment Simulator	Essery et al. (2003), Best (2005), Best et al. (2006)	4	2
LUMPS	Local-scale Urban Meteorological Parameterization Scheme	Grimmond and Oke (2002), Offerle et al. (2003)	2	1
NKUA	University of Athens Model	Dandou et al. (2005)	1	1
MORUSES	Met Office Reading Urban Surface Exchange Scheme	Harman et al. (2004a,b), Porson et al. (2009, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*)	2	1
MUCM	Multilayer Urban Canopy Model	Kondo and Liu (1998), Kondo et al. (2005)	1	1
NJU-UCM-S	Nanjing University-Urban Canopy Model-single layer	Masson et al. (2000), Kusaka et al. (2001)	1	1
NJUC-UM-M	Nanjing University-Urban Canopy Model-multiple layer	Kondo et al. (2005), Kanda et al. (2005a,b)	1	1
NSLUCM/NSLUCMK/ NSLUCM-WRF	Noah land surface model/Single-layer Urban Canopy Model	Kusaka et al. (2001), Chen et al. (2004)	3	3
SM2U	Soil Model for Submesoscales (Urbanized)	Dupont and Mestayer (2006), Dupont et al. (2006)	1	1
SNUUCM	Seoul National University Urban Canopy Model	Ryu et al. (2009)	1	1
SRUM2/SRUM4	Single Column Reading Urban Model tile version	Harman and Belcher (2006), Porson et al. (2009)	4	1
SUEB	Slab Urban Energy Balance Model	Fortuniak et al. (2004, 2005)	1	1
SUMM	Simple Urban Energy Balance Model for Mesoscale Simulation	Kanda et al. (2005a,b; 2007), Kawai et al. (2007, 2009)	1	1
TEB	Town Energy Balance	Masson (2000), Masson et al. (2002), Lemonsu et al. (2004)	1	1
TEB07	Town Energy Balance 7	Hamdi and Masson (2008)	1	1
TUF2D	Temperatures of Urban Facets 2D	Krayenhoff and Voogt (2007)	1	1
TUF3D	Temperatures of Urban Facets 3D	Krayenhoff and Voogt (2007)	1	1
VUCM	Vegetated Urban Canopy Model	Lee and Park (2008)	1	1

(a) Class details

Class code	Cap	VL92
Fluxes included (F)		
All fluxes (a)	14	10
No Q_E (e)	6	2
No Q_F (f)	13	16
Neither Q_E nor Q_F (g)	0	5
1 Vegetation (V)		
Not included (n)	9	14
Separate tile (s)	18	13
Integrated (i)	6	6
2 Q_F (AN)		
Negligible or ignored (n)	12	26
Prescribed (p)*	6	1
Internal Temp. (i)*	4	4
Modelled (m)*	5	1
i,p*	6	1
3 Temporal Q_E variation (T)		
None (n)	13	26
Fixed (f)	6	2
Variable (v)	14	5
4 Urban Morphology (L)		
Bulk (L1)	8	6
Single layer (L2)	5	7
(L3)	13	13
Multiple layer (Lm) (L4)*	2	2
(L5)*	1	1
(L6)*	1	1
(L7)*	3	3
5 Facets & orientation (FO)		
Bulk (1)	5	5
No orientation (n)	12	16
Orientation no intersec. (o)	10	6
Orientation & intersect. (i)	6	6
6 Reflection (R)		
Single (1)	11	11
Multiple (m)	15	15
Infinite (i)	7	7
7 Albedo, Emissisivity (AE)		
Bulk (1)	5	6
Two facets (2)	6	4
Three or more facets (3)	22	23
8 ΔQ_S (S)		
Residual (r)*	6	7
Conduction (c)	24	23
Net radiation based (n)*	3	3

(b) Turbulent Flux Methods

	Q_H B	Q_H V	Q_E B	Q_E V
Resistance/Conductance (G)				
Bulk (1)	9	11	9	11
Single layer (3)	19	4	12	4
Multilayer (4)	5	2	3	2
Surface temp/moisture (Z)				
Bulk (1)	9	8	9	8
Single layer facet (3)	14	3	9	3
Multi layer facet (4)	8	4	4	3
Penman/comb no surface (P)	2	2	2	3
Air temperature/moisture (A)				
Single layer air (3)	24	14	17	14
Multi layer air (4)	6	2	4	2
Forcing height (F)	3	1	3	1
Other (O)				
Not included (N)	-	14	7	14
Specified/fixed evap. rate (E)	-	-	2†	2§
Soil only (W)	-	2	-	2
Human perspiration (H)	-	-	1	-

(c) Urban morphology (L) and vegetation (V) combinations (see b)

	B Q_H	V Q_H	B Q_E	V Q_E
L1	113 (4)	113 (2)	113 (2)	113 (2)
		N (2)	113 (2)	N (2)
	1P3 (2)	1P3 (2)	1P3 (2)	1P3 (2)
L2	113 (2)	113 (2)	113 (2)	113 (2)
	11F (1)	11F (1)	11F (1)	11F (1)
	31F (2)	N (2)	31F (2)	N (2)
	333 (2)	N (2)	N (2)	N (2)
L3	333 (9)	113 (3)	333 (3)	113 (3)
		133 (1)	333 (1)	133 (1)
		343 (2)	333 (1)	343 (1)
			343E (1)	3P3E (1)
		N (3)	333 (1)	N (1)
			N (2)	N (2)
	334 (1)	N (1)	334 (1)	N (1)
	343 (3)	343 (2)	343 (1)	343E (1)
			E (1)	343 (1)
		N (1)	343 (1)	N (1)
L4	444 (2)	W (2)	N (1)	W (1)
			H (1)	W (1)
L5	434 (1)	434 (1)	434 (1)	434 (1)
L6	444 (1)	N (1)	444 (1)	N (1)
L7	343 (2)	N (2)	N (2)	N (2)
	434 (1)	434 (1)	434 (1)	434 (1)

	B Q_H	V Q_H	B Q_E	V Q_E
Vn	113 (2)	N (2)	113 (2)	N (2)
	31F (2)	N (2)	31F (2)	N (2)
	333 (5)	N (5)	333 (1)	N (1)
			N (4)	N (4)
	334 (1)	N (1)	334 (1)	N (1)
	343 (3)	N (3)	343 (1)	N (1)
			N (2)	N (2)
	444 (1)	N (1)	444 (1)	N (1)
Vs	113 (4)	113 (4)	113 (4)	113 (4)
	11F (1)	11F (1)	11F (1)	11F (1)
	333 (4)	113 (1)	333 (1)	113 (1)
		133 (1)	333 (1)	133 (1)
		343 (2)	333 (1)	343 (1)
			343E (1)	3P3E (1)
	343 (1)	343 (1)	E (1)	343 (1)
	434 (1)	434 (1)	434 (1)	434 (1)
	444 (2)	W (2)	N (1)	W (1)
			H (1)	W (1)
Vi	1P3 (2)	1P3 (2)	1P3 (2)	1P3 (2)
	333 (2)	113 (2)	333 (2)	113 (2)
	343 (1)	343 (1)	343 (1)	343E (1)
	434 (1)	434 (1)	434 (1)	434 (1)

(d) Energy Balance Closure

		Yes	No		VL92 actual closure	
Forced		24	9		Yes	No
Closure forced by	Surface Temperature	6	Model checks closure			
	ΔQ_S residual	17	Yes	No	31	2
	Turbulent fluxes available energy limited	1	7	2		

FIG. 2. Characteristics used to classify the models (see Fig. 1). (a) Class code and number of models that have this capability (cap) and were applied this way for VL92. Classes with few models (asterisks) are amalgamated for analysis. (b) Approaches used to simulate the built (B) and vegetated (V) turbulent heat fluxes (Q_H, Q_E). The numbers are for the VL92 runs. (c) Combined features [from (b)] GZA or "Other" used in the turbulent flux modeling for the VL92 runs with the numbers in each class shown. (d) Energy balance closure approach. In (b), the dagger symbol indicates that one is also 343 and § indicates that one is 3P3 and one is 343.

to ensure adequate representation. In the tile approach, the built and vegetated fluxes are typically weighted by their respective plan area fractions to contribute to total fluxes (e.g., Lemonsu et al. 2004). The integrated case (Vi) is the most physically realistic as it allows for direct interaction of built and vegetated surfaces. This additional complexity may require increased computing resources and parameter values.

b. *Anthropogenic heat fluxes ("AN" in Figs. 1 and 2)*

The magnitude of Q_F varies across a city. Typically it will be greatest in the densest part of the city (Oke 1988; Grimmond 1992; Ichinose et al. 1999). But even low absolute Q_F values may be important where they exceed the radiative forcing (e.g., cloudy, cold winters with low solar input).

Similar to Q_E, not all models consider Q_F. The four general approaches are

1) ANn: Q_F is assumed to be zero, negligible, or ignored;
2) ANp: Q_F is assumed to be a fixed amount that is required as specified input to the model, or is directly coded into the program;
3) ANi: Q_F is calculated based on assumed internal building temperature; and
4) ANm: Q_F is calculated and incorporates internal heat sources from buildings, and/or mobile sources associated with traffic, and/or metabolism.

Models that calculate Q_F typically include the heat related to internal heating of buildings as a minimum. A fixed temperature is assigned internally and this may, or may not, be allowed to vary seasonally or diurnally. Alternatively a fixed minimum (maximum) temperature is used so the internal temperature of the building may vary but within limits. The heat flux from traffic typically is based on assumptions about traffic flow from vehicle counts. The models that calculate Q_F in more detail, using a building energy model, mostly use the method of Kikegawa et al. (2003).

Beyond the internal temperature, the introduction of Q_F requires consideration of where the heat is released or added to the atmosphere; for example, whether heat is added within or above the canyon.

c. *Anthropogenic heat fluxes: Temporal variation ("T" in Figs. 1 and 2)*

Anthropogenic heat flux Q_F varies both diurnally and seasonally (e.g., Sailor and Lu 2004; Offerle et al. 2005; Pigeon et al. 2007; Lee et al. 2009), although only some models consider this. Models that prescribe a fixed value (Tf) are likely to provide too much Q_F at night and insufficient quantities in the day; they will also not capture peak values normally associated with commuting (seasonally this peak may be associated with low solar radiation forcing). The inclusion of a diurnal and/or seasonal cycle (Tv) is more significant for certain applications when the modeled fluxes must be correct for specific short time periods. It is less significant when applications are not concerned with diurnal patterns.

d. *Urban morphology ("L" in Figs. 1 and 2)*

Urban morphology affects radiative and turbulent heat exchanges. A number of approaches are used to capture these features, including

- L1: Slab or bulk surface;
- L2: single-layer approaches, which separate the surface into a roof and canyon (wall plus road) or
- L3: where the three facets (roof, wall, and road) are treated separately; and
- L4–L7: multiple-layer approaches, which divide one or more of the facets into layers or patches.

Slab models represent the urban form as a flat horizontal surface with appropriate "bulk" radiative, aerodynamic, and thermal characteristics. This has the advantage of simplicity and reduced computational time and parameter requirements. Single-layer models simplify the urban form to an urban canyon with a roof, wall, and a road. This allows for more realistic representations of radiative trapping and turbulent exchange (Masson 2000; Kusaka et al. 2001; Harman et al. 2004a,b; Lee and Park 2008). Parameter values are assigned for each facet and one set of energy exchanges per facet is modeled. Multilayer schemes divide the walls into a number of vertical and/or the roof and road into a number of horizontal patches; each with their own parameter values and energy exchanges modeled. For some models this allows for variable building height, and for others even differing roof, wall, and road characteristics. Note that the range of multilayer models L4–L7 is not exhaustive; rather it covers the range compared here.

e. *Urban morphology: Facets and orientation ("FO" in Figs. 1 and 2)*

Models can be further subdivided by how urban canyon morphology, specifically the number of facets and orientations, are dealt with. Models include those that assume no facets (or orientation) and hence a bulk (or slab) surface (FO1), those that assume one infinitely long canyon (FOn), and those that have infinitely long canyons that run in two cardinal directions (FOo). The canyons may be fixed in orientation and neglect shading

or assume a random distribution of street canyons within the domain. Alternatively, the canyon may be modeled assuming two walls that have sunlit and shaded fractions that vary through the day and year. More realistic models also include an intersection between canyons (FOi), significantly increasing the number of the interactions with other facets that need to be computed.

f. Radiative fluxes: Reflections ("R" in Figs. 1 and 2)

As $K\downarrow$ and $L\downarrow$ are provided, it is the $K\uparrow$ and $L\uparrow$ that are modeled. Beyond the morphology, and therefore the degree of detail needed for the surface parameters, the major differences relate to the number of reflections assumed—R1: single; Rm: multiple; Ri: infinite.

The single reflection model is the least computationally intensive and used in both slab and single-layer models. Models that simulate multiple reflections include both single-layer and multiple-layer models. Infinite reflections may be accounted for by slab, single-layer, and multilayer models.

For longwave radiation, slab models determine one surface temperature, whereas for facet-specific models, multiple surface temperatures are calculated (Figs. 2b,c). The surface temperatures then provide the forcing for Q_H and ΔQ_S.

g. Radiative fluxes: Albedo and emissivity ("AE" in Figs. 1 and 2)

The albedo and emissivity values that determine the radiative fluxes may either be defined as a single value (bulk, AE1); as two facets, similarly to the L2 category (AE2); or may consist of combinations of various facets, analogous to the L3 (or greater) category of models (AEf).

h. Storage heat flux (ΔQ_S; "S" in Figs. 1 and 2)

The ΔQ_S is significant in urban areas given the materials and morphology of the urban surface (Grimmond and Oke 1999a). In urban models, it is determined in the following ways:

1) Sr: difference or residual of the energy balance,
2) Sc: solution of the heat conduction equation by dividing the facets into a number of thickness layers, and
3) Sn: function of Q^* and surface characteristics.

All three methods are used by slab or bulk models (Fig. 3). For all three, the ability to model the outgoing longwave radiation will impact the values obtained given the common need to model surface temperature.

For those models in which heat storage is calculated as the residual of the surface energy balance, assumptions as to which fluxes are included (specifically Q_F and Q_E) are important. The second method, the solution of the heat conduction equation, is used extensively by slab, single-, and multiple-layer models. It requires various parameters for each (sub) facet, including: number of layers, layer thickness, thermal conductivity, and volumetric heat capacity of the various layer materials (Table 2). The number of layers resolved varies between 1 and 48, and may be of fixed or variable thickness. Currently, none account for changing water content of built materials associated with rain, so the material parameters are static. Some solve the heat conduction equation using the force-restore method, while others solve the one-dimensional heat conduction equation.

The third approach is to use a fraction of Q^*(Sn). Some models take into account the diurnal pattern of the flux through the objective hysteresis model (Grimmond et al. 1991).

i. Other features

The following characteristics are not explicitly used to classify the models in this evaluation but are presented here because of differences between models. They do not necessarily result in the models being grouped differently to the classifications above; that is, models fall into some common groupings across model classes (Figs. 2, 3).

1) TURBULENT SENSIBLE (Q_H) AND LATENT HEAT (Q_E) FLUX

Typically surface resistance (or its inverse, conductance) schemes are used to model Q_H and Q_E ("G" in Fig. 1, Fig. 2b). Depending on urban morphology, these consist of either single (G3) or multiple (G4) resistance networks, which account for the number of facets and layers that are resolved. Bulk models (G1) have the simplest resistance network (Figs. 1, 2b). A wide range of resistance schemes is used (e.g., Rowley et al. 1930; Clarke 1985; Zilitinkevich 1995; Guilloteau 1998; Harman et al. 2004b). To determine the resistance the wind profile within/above the canyon, roughness length, and displacement length or drag coefficients and atmospheric stability may be taken into account. Drag is either not considered or is calculated using roughness length, exponential wind profile, or distributed drag. Exchange between the canopy air and building surfaces may be parameterized by a roughness length approach or distributed sources of heat (generally in conjunction with a distributed drag approach).

The number of temperatures resolved, which drive the gradients, varies both for the surface and the air (within the canyon; "Z" and "A" respectively in Figs. 1 and 2b), and these are related to morphology and the number

	Vn	Vs	Vi	ANn	ANp	ANi	ANm	ANip	Tn	Tf	Tv	L1	L2	L3	L4	L5	L6	L7	FO1	FOn	FOo	FOi	R1	Rm	Ri	AE1	AE2	AEf	Sr	Sc	Sn	Capability
Vn				1		2	2	4	1	4	4	1	2	3			1	2	1		6	2	1	8		1	2	6		9		Vn
Vs				9	4	1	2	2	9	2	7	5	3	6	3			1	2	10	4	2	8	3	7	2	4	12	5	12	1	Vs
Vl				2	1	1	2		2		4	2		3		1			2	2		2	2	4		2		4	1	3	2	Vl
ANn	10	11	5						12			7	2	1	1			1	5	5		2	8	2	2	5	4	3	1	9	2	ANn
ANp	1										5		1	4						3	2		2	2	1			5	2	3		ANp
ANi	2	1	1							1	3				1	1		2			2	2		3	1			4	1	3		ANi
ANm		1									6	1		3	1		1		2	2	2	2	1	4	1			6	2	3	1	ANm
ANip	1									5	1		2	4					2	4				4	2		2	4		6		ANip
Tn	10	11	5	26								7	2	1	1			1	5	5		2	8	2	2	5	4	3	1	9	2	Tn
Tf	1	1			1	1							2	3	1					1	5			4	2		2	4	1	5		Tf
Tv	3	1	1			3	1	1				1	1	8	1	1	1	2		6	5	4	3	9	3			15	4	10	1	Tv
L1	2	2	2	5	1				5	1									5	3			6		2	5	2	1		5	3	L1
L2	4	3		7					7											3	2		2	2	1		4	1	1	4		L2
L3	5	5	3	12				1	12		1									6	5	1	3	7	2			12	2	10		L3
L4		2				1	1			1	1									2	1			1	2			3		2	1	L4
L5			1			1					1										1			1				1		1		L5
L6	1			1					1												1			1				1		1		L6
L7	2	1		1		2			1		2									1	2			3				3	1	2		L7
FO1	1	2	2	5					5			5											5			5				3	2	FO1
FOn	9	5	2	14	1			1	14	1	1	1	7	8									5	2	5		4	8	1	10	1	FOn
FOo	2	4		3		2	1		3	1	2			3	2			1					1	7	2		2	8	3	7		FOo
FOi	2	2	2	4		2			4		2			2		1	1	2						6				6	2	4		FOi
R1	2	7	3	10	1				10	1		6	2	3					5	5	1					5	2	4		8	3	R1
Rm	8	3	3	12		3			12		3		2	8		1	1	3		6	3	6					2	13	3	12		Rm
Ri	4	3		4		1	1	1	4	1	2		3	2	2					5	2						2	5	3	4		Ri
AE1	2	2	2	5	1				5	1		6							5	1			6							3	2	AE1
AE2	2	2		4					4				4							4			2	2						6		AE2
AEf	10	9	4	17		4	1	1	17	1	5		3	13	2	1	1	3		11	6	6	3	13	7				6	15	1	AEf
Sr		5	2	5		1	1		5	1	1		1	3	2			1		2	3	2		4	3			7				Sr
Sc	13	8	2	19		3		1	19		4	3	6	10		1	1	2	3	13	3	4	8	11	4	3	4	16				Sc
Sn	1		2	2	1				2	1		3							2	1			3			3						Sn
VL92 use	Vn	Vs	Vi	ANn	ANp	ANi	ANm	ANip	Tn	Tf	Tv	L1	L2	L3	L4	L5	L6	L7	FO1	FOn	FOo	FOi	Rs	Rm	Ri	AE1	AE2	AEf	Sr	Sc	Sn	

FIG. 3. Analysis of the individual characteristics (Fig. 2) gives this combination of approaches: (top right) the capability of the models and (bottom left) how the models were applied for the VL92 runs analyzed here. See Figs. 1 and 2 for explanation of class codes. For example, with respect to the capability of the 33 models: five that model vegetation in a separate tile (Vs), have slab morphology (L1); while in application to the VL92 dataset, only two were run this way.

TABLE 2. Urban parameters used by the number of models indicated for the VL92 runs and also those indicated as being applicable for the models. Each subscript refers to a separate parameter: f = roof; r = road; w = wall; v = pervious; t = tree; H = building; g = grass, s = soil, m = momentum; h = heat, u = urban; b = bulk. For some models § indicates state variables, a dagger indicates fundamental parameters, or an asterisk indicates derived parameters. Also noted is visible (VIS), ultraviolet (UV), and near infrared (NIR).

Name (units)	Parameter	No. models (VL92)	No. models (capability)
Radiative			
Albedo (-)	$\alpha_{f,r,w,v,t,g,b}$	24, 22, 19, 9, 6, 6, 4	24, 22, 19, 9, 6, 6, 4
VIS, UV, and NIR albedo of vegetation (-)	$\alpha_{\text{VIS,UV,NIR}}$	1, 0, 3	2, 1, 4
Emissivity (-)	$\varepsilon_{f,r,w,v,t,g,b}$	23, 21, 19, 5, 6, 5, 2	23, 21, 19, 5, 6, 5, -
Roughness			
Roughness length above canyon (m)	z0m,c; z0h,c	10, 8	10, 10
Material roughness length for heat (m)	$\text{z0h(mat)}_{f,r,w}$	14, 16, 6	4, 16, 6
Material roughness length for momentum (m)	$\text{z0m(mat)}_{f,r,w}$	16, 14, 6	17, 14, 6
Effective roughness length for heat (m)	$\text{z0h(eff)}_{f,r,w}$	—	5, 0, 2
Effective roughness length for momentum (m)	$\text{z0m(eff)}_{r,w}$	—	5, 2
Roughness length of grass/tree surfaces (m)	$\text{z0m}_{g,t}$	4, 5	—
Bulk roughness length (m)	z0b	1	2
Sublayer Stanton no. (-)	S_n	1	1
Zero-place displacement height (m)	$\text{d0}_{f,r,w}$	2, 1, 1	2, 1, 1
Thermal characteristics			
Volumetric heat capacity (MJ K^{-1} m^{-3})	$c_{\text{Pf},r,w,v,t,g,s}$	23, 24, 20, 8, 6, 5, 12	23, 24, 20, 8, 6, 5, 12
Thermal conductivity (MJ K^{-1} m^{-3})	$K_{f,r,w,v}$	22, 22, 21, 8	22, 22, 21, 8
No. of layers (-)	$\text{nl}_{f,r,w,v,s}$	19, 18, 17, 3, 14	19, 18, 17, 3, 14
No. of walls modeled (-)	nw	17	17
Layer thickness (m)	$\text{dl}_{f,r,w,v,s}$	22, 20, 2, 13	22, 20, 2, 16
Urban morphology			
Height of measurements/reference height (m)	Z_{ref}	15	16
Mean height (m)	$Z_{H\dagger,g,t,\text{all}}$	17, 4, 5, 5	17, 4, 6, 7
Avg building separation/canyon width (m)	$W_{X\dagger}$	8	8
Avg width of buildings (m)	$L_{X\dagger}$	8	8
Vertical floor density (building proportion > n stories) (-)	λ_B	3	3
Sky view factor (-)	SVF	1	1
Mean block length (m)	$L_{Y\dagger}$	3	3
Mean long axis azimuth of walls (m)	az	2	2
Canyon height-to-width ratio (-)	λ_s^*	2	11
Frontal area index (-)	λ_F^*	6	6
Wall to nonbuilt horizontal area (-)	λ_{wnb}^*	2	2
Plan area			
Fraction of area (-)	$\lambda_{f,r,v,g,t}^*$	19, 19, 10, 8, 6	19, 19, 10, 8, 6
Anthropogenic heat flux			
Total anthropogenic heat flux (W m^{-2})	$Q_{f,\text{tot}}$	6	6
Sensible heat flux from vehicles (W m^{-2})	$Q_{H,\text{traf}}$	4	4
Sensible heat flux from industry (W m^{-2})	$Q_{H,\text{ind}}$	3	3
Temperature §			
Mean internal building temperature (K)	T_{int}	6	8
Deep temperature (K)	$T_{\text{dr},s}$	6, 6	6, 8
Facet temperature (K)	$T_{f,r,w,s}$	6, 6, 5, 4	6, 6, 5, 6
Min; max internal building temperature (K)	$T_{\text{int,min;max}}$	6, 3	6, 3
Vegetation/soil specific			
Vegetation wilting point (m^3 m^{-3})	S_{wilt}	9	10
Rooting depth of grass/tree (m)	$\text{dr}_{g,t,v}$	3, 3, 2	4, 4, 3
Min stomatal resistance (s m^{-1})	R_{min}	5	5
Leaf area index of vegetation within the urban canyon (-)	LAI	6	7
Vegetation thermal inertia (J m^{-2} K^{-1} $s^{-1/2}$)	T_i	1	1
Parameters used in radiation stress function (-)	R_s	3	3
Parameter used in vapor pressure deficit function (-)	P_v	3	3
B parameter (-)	B	12	12
Sand/clay/loam/quartz content of soil (-)	$S_{s,c,l,q}$	6, 4, 3, 1	7, 7, 4, 2
Max vegetation canopy water holding capacity (mm)	$M_{t,g}$	6, 3	6, 4
Optimum temperature in temperature stress function (K)	T_{opt}	1	1
Coef for max interception water storage capacity (-)	I_s	1	2

TABLE 2. (*Continued*)

Name (units)	Parameter	No. models (VL92)	No. models (capability)
Ecosystem respiration parameter (-)	E_r	1	2
Ratio d(biomass)/d(LAI) (-)	d/d	1	2
e-folding time for senescence (-)	E_f	1	2
Cuticular conductance (-)	C_c	1	2
Max air saturation deficit (100 g kg^{-1})	A_s	1	2
Leaf area ratio sensitivity to nitrogen (m^2 kg^{-1})	S_N	1	2
Lethal min value of LA ratio (m^2 kg^{-1})	L_{la}	1	2
Nitrogen concentration of biomass (m^2 kg^{-1})	N_c	1	2
Root fraction (%)	R_f	1	2
Tree coverage (%)	T_c	1	1
Sunny spots (%)	S_s	1	1
Canopy solar absorptivity (-)	C_a	1	1
Canopy solar transmissivity (-)	C_m	1	1
Canopy thermal time constant (-)	C_t	1	1
Tree evaporative cooling coef (-)	E_c	1	1
Moisture availability			
Moisture availability (m^3 m^{-3})	$M_{f,r,s}$	3, 3, 7	4, 4, 8
Hydraulic conductivity of the soil at saturation (m s^{-1})	K_{sat}	13	13
Critical normalized soil water content for stress (m s^{-1})	S_{crit}	2	3
Air dry soil moisture content limit (m s^{-1})	S_{lim}	3	3
Soil suction experienced in the soil at saturation (m)	S_{suc}	9	10
Max soil moisture content (field capacity; m^3 m^{-3})	$M_{s,\text{max}}$	12	12

of vertical layers in the model. Many assume Monin–Obukhov similarity holds, which may not be applicable within the urban canyon or within the roughness sublayer (Roth 2000). However, given the lack of well-tested alternatives, currently, this may be the most appropriate approach.

As Q_H is calculated typically using surface temperature to force the gradient, a balance is inherent in the solution of surface temperature between the $L\uparrow$, Q_H, and heat conduction. Depending on the model objective, performance may be improved for one flux at the expense of another. Models that use a combination method (P, Penman–Monteith, or combination-type approach) do not need to determine the surface temperature to calculate Q_H, but still need to allow for the transport of heat away from the surface.

The approaches taken to model resistances (G), surface temperature (Z), and air temperature (A) result in a large number of combinations (Fig. 2c, expressed in GZA order). Here they are shown relative to the urban morphology classes (L1–L7; Fig. 1) and the vegetation class (Vn, Vs, Vi). The approach taken for each turbulent flux (Q_H, Q_E) for the built (B) and vegetated (V) part of the surface are shown. It is clear that the earlier classifications (Fig. 2a) do not produce common characteristics for these fluxes. Given the wide range of approaches, these are not investigated in further detail in this paper. Subsequent analysis of a larger dataset will investigate this. For the calculation of Q_H for the built (B) fraction of the surface, the two most common classes of the nine different combinations are

1) 333: single-layer resistance (G3), surface temperature (Z3), and air temperature (A3); and
2) 113: bulk resistance per facet (G1), surface temperature (Z1), and a single air temperature (A3).

For the vegetated surfaces the two most common classes for Q_H are N: Q_H is not calculated; and 113. For Q_E from built surfaces the predominant classes are N, 113, and 333; but also of note are those models that account for the evaporative loss of water in one time step immediately following precipitation with a fixed rate of evaporation (E). For Q_E from vegetated surfaces, the predominant classes are also N and 113. Two models, which do simulate Q_H and Q_E for vegetated area, account for evaporation from soil moisture only and not the loss of water through vegetation. In these cases the soil temperature and moisture profile are calculated using the approach of Tremback and Kessler (1985). In urban areas bare soil is rare, some sort of vegetation is most likely to be present.

2) ENERGY BALANCE CLOSURE

Not all models explicitly force or check that they have energy balance closure [i.e., that Eq. (1) holds; Fig. 2d]. Lack of closure may result from numerical instabilities or lack of precision in the code, from a lack of evaluation, or

from inconsistent assumptions. Closure may be forced in a number of ways: through the calculation of ΔQ_S at the end of each time step as a residual, by updating the surface temperature of the facets, or by restricting the turbulent heat fluxes to the available energy ($Q^* - \Delta Q_S$). Closure is an important issue when the land-surface scheme is part of a long-term climate model simulation; without it, there may be long-term bias in the model.

3) ANTHROPOGENIC WATER FLUX AND OTHER CAPABILITIES

Water can be added to the urban environment by human activity. Water is released by combustion processes, cooling towers, and by people, which is equivalent to the Q_F release (anthropogenic latent heat flux). One model takes into account the loss of water through perspiration (a source of Q_E). Given there are very few estimates of this term (Heiple and Sailor 2008; Moriwaki et al. 2008), and it is likely to be small in many settings, it is not surprising that it has not been included in most models. The term may be important in very dry areas (e.g., high-latitude cities in winter, hot dry cities) and in areas with excessive air conditioning. The second significant source of water comes via the pipe network, most typically as irrigation (e.g., garden sprinkling) or broken water pipes. In many suburban areas, if gardening is a common residential activity, this can be a large additional source of water relative to precipitation, especially during the summertime (e.g., Grimmond and Oke 1986; Grimmond et al. 1996). Estimating this component requires assumptions in the algorithms and/or the input data to define: 1) how much, and when, water is applied to the area and 2) where in the area it is released (e.g., to all vegetated surfaces or just to irrigated grass). The representation of this source is important (e.g., Mitchell et al. 2001) but has been considered in few models (e.g., Grimmond and Oke 1991).

The presence of snow cover will influence the energy balance of urban regions, affecting the albedo and, during periods of snowmelt, acting as a significant sink for latent energy (Lemonsu et al. 2008). For models with facets, the energy budgets of horizontal surfaces (roof and road) will be the most significantly affected, with additional energy budgets for these surfaces being necessary (Masson 2000).

j. Model uniqueness

Using the 31 individual characteristics to classify the models compared (Fig. 2a), 26 unique combinations occur (Fig. 3). This varies between model capability and actual use (demonstrated here for a dataset termed VL92; see section 4). For example, 21 models have the capability to account for Q_F but only 7 utilize this capability for the VL92 application. Although there are preferred approaches (e.g., for Q_F Tv over Tf), there is a notable diversity; models that have a similar approach for one aspect frequently use quite different approaches for other model components.

3. Model inputs

Inputs of three general types are required to model urban areas: 1) site parameters to describe the surface morphology and materials; 2) time series of atmospheric or forcing variables as boundary conditions; and 3) initial thermodynamic and moisture state conditions. The complexity of urban areas and diversity of surface description methods in the 33 models results in more than 145 (or >200 if individual layer values are considered) different parameters and state variables being needed for all of the models. The parameters fall into nine broad classes (Table 2). Some parameters, which are unique to individual models, can be derived from more basic parameters (Table 3; Fig. 4). Given the large effort needed to collect these data over the wide range of urban areas globally, or even within individual countries (e.g., Feddema et al. 2005; Ching et al. 2009), we encourage model developers to use common parameters. Also, it is important that the parameters are clearly defined and not open to misinterpretation (Loridan et al. 2010).

Morphometric parameters vary greatly, using either basic information (e.g., height, width) from which required parameters are calculated (e.g., canyon aspect ratio, sky view factor), or the "higher" level parameters as the inputs. Table 2 lists basic parameters from which higher-level parameters can be calculated.

Urban material related parameters are required to account for radiative transfer (e.g., albedo, emissivity) and thermal characteristics. Because of the different ways to describe the surface (Figs. 1, 2), there are varying numbers of models that use particular parameters (Table 2). All models use some form of albedo but this may be a single bulk albedo (α_b), or albedos for the roof (α_f), wall (α_w), road (α_e) and so on. Thermal properties are specified explicitly either relative to mass (specific heat capacity) or volume (volumetric heat capacity), or implicitly from model lookup tables.

As noted in Fig. 2, Q_F is dealt with in a variety of ways. For those using a fixed value, a model parameter has to be specified or alternatively, internal building temperature may need to be specified (in Table 2, it is included under temperatures but could be specified under Q_F).

Temperatures are required for many models. These may be initial state conditions (e.g., facet temperatures), which will evolve during the run, or require model spinup of sufficient time, or may be fixed for the duration of the run (e.g., deep soil temperature). In many applications, it

TABLE 3. Fundamental morphometric parameters (units of meters) that can be used to derive dimensionless morphometric parameters. Also shown are thermal parameters. Note many different names are used for the same parameters. Refer to Fig. 4 for further definitions.

Parameteralternative names	Symbol	Derivation
Fundamental morphometric parameters (m)		
Mean building height	z_H	
Mean building length Mean building width	$L_{\overline{XY}}$	$L_{\overline{XY}} = \frac{\sum_{i=1}^{n}(L_{Xi} + L_{Yi})}{2n} = \sqrt{L_{\overline{X}} L_{\overline{Y}}}$ n = all relevant directions for weighting
Mean canyon width Mean building separation	$W_{\overline{XY}}$	$W_{\overline{XY}} = \frac{\sum_{i=1}^{n}(W_{Xi} + W_{Yi})}{2n} = \sqrt{W_{\overline{X}} W_{\overline{Y}}}$
Mean block length Mean canyon length	$D_{\overline{XY}}$	$D_{\overline{XY}} = L_{\overline{XY}} + W_{\overline{XY}}$
Derived morphological parameters		
Canyon height-to-width ratio Canyon aspect ratio	λ_s	$\lambda_S = \frac{z_H}{W_{\overline{XY}}}$
Plan area ratio Roof area ratio Building coverage ratio Building fraction	λ_P	$\lambda_P = \frac{L^2_{\overline{XY}}}{D^2_{\overline{XY}}}$
Road area fraction	λ_r	$\lambda_r = \frac{W_X D_Y + W_Y L_X}{D^2_{\overline{XY}}}$
Frontal area index Building frontal density	λ_F	$\lambda_F = \frac{z_H L_{\overline{XY}}}{D^2_{\overline{XY}}}$
Wall to nonbuilt horizontal area	λ_{WNB}	$\lambda_{\mathrm{WNB}} = \frac{2(L_X + L_Y) z_H}{W_X D_Y + W_Y L_X}$
Thermal parameters: fundamental		
Density (kg m^{-3})	ρ	
Specific heat (J kg^{-1} K^{-1})	c	
Thermal conductivity (W m^{-1} K^{-1})	λ	
Thermal parameter: derived		
Volumetric heat capacity (J m^{-3} K^{-1})	$C = \rho c$	

is likely to be difficult to have realistic or observed values to meet the need for the temperature profile within a building or the soil to be prescribed. This may mean that some models require a long initialization period (spinup) to ensure that the temperature profiles are stable and representative of expected conditions.

For the models that use a vegetation tile, all the parameters required are not summarized in Table 2.

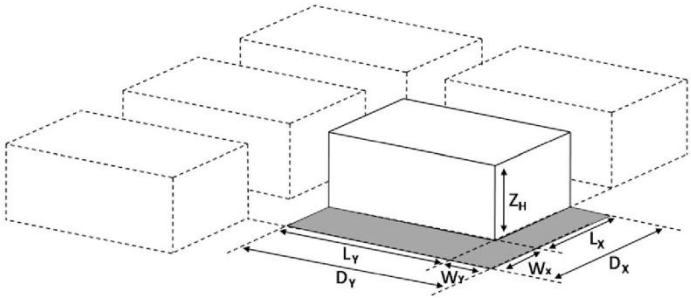

FIG. 4. Definitions of fundamental morphometric parameters (see Table 3). Note that with changing wind direction the frontal area index will vary.

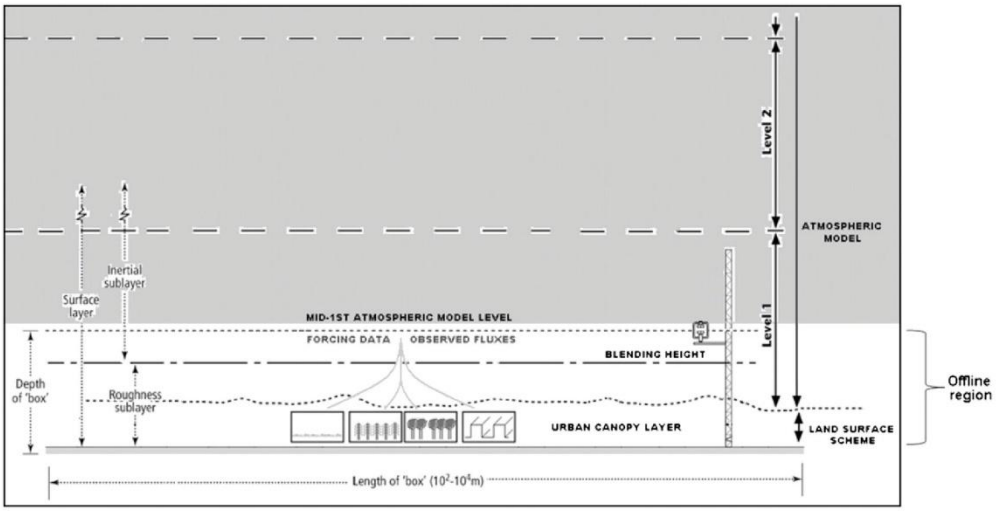

FIG. 5. Urban land-surface schemes simulate exchanges between the urban surface and the first layer of larger-scale atmospheric models. The observed fluxes and the forcing data are representative of the same level since they are above the roughness sublayer or blending height.

Parameter values, based on class selection, have been determined for extensive nonurban vegetated areas, and are assigned through model lookup tables. Model users have selected the vegetation class (e.g., grassland, deciduous or evergreen woodland, and/or bare soil) that they think is most appropriate in relation to the urban region they are modeling.

Soil moisture characteristics require both initial values and fixed parameters. These state variables have similar constraints and implications to that of the temperature. As urban areas often have disturbed soils and additional materials mixed into the media, it may mean that adoption of rural soil physical properties for parameters is not appropriate.

4. The International Urban Energy Balance Model Comparison Project

The methodology adopted here follows that of PILPS (Henderson-Sellers et al. 1993), which provided insight into both the models and real world processes. This allows the relative importance of key parameters to be determined and an assessment of the level of complexity required to produce reliable results. The International Urban Surface Energy Balance Model Comparison Project has been endorsed by the Global Energy and Water Cycle Experiment (GEWEX) Global Land–Atmosphere System Study (GLASS) and World Meteorological Organization Expert Team on Urban and Building Climatology.

The procedure for the comparison requires individual modeling groups (users and/or developers) to run their model(s) offline using forcing data provided for the top of the model, as would be provided by an atmospheric model (Fig. 5). This implies that parameter values should be representative of the observational footprint (see discussion in Masson et al. 2002). There is no feedback to larger-scale conditions within the modeling domain, so no larger-scale advection can occur, as would be present in a mesoscale or larger-scale model. The temporal resolution of analysis is typically 30 or 60 min, but individual models may be run at higher temporal resolution (1.5–300 s) and then average or sample their data back to the specified time interval of analysis (60 min). The spatial scale for both the measurements and models is the local or neighborhood scale (10^2–10^4 m). However there is no actual grid size because the models are run in single column mode. The observed fluxes and the forcing data are taken from tall towers that have the sensors located above the roughness sublayer (Grimmond and Oke 1999b; Roth 2000; Masson et al. 2002; Grimmond et al. 2004; Grimmond 2006). This height is equivalent to being above or at the blending height and is typically taken as the first atmospheric layer in mesoscale or larger-scale models (Fig. 5). The rationale for offline simulation is that although larger-scale circulation models may be accurate at the macroscale, their outputs will often be incompatible with those required as inputs to mesoscale urban surface models (Pitman et al. 1990). Equally, running such models

offline prevents feedbacks between climate and land surface, meaning that the sensitivity of the land-surface schemes themselves can be examined while the overlying atmospheric conditions are effectively held fixed (Wilson et al. 1987; Henderson-Sellers and Dickinson 1992).

The forcing data provided to participants were collected from a light industrial site in Vancouver, British Columbia, Canada (termed here VL92; Voogt and Grimmond 2000; Grimmond and Oke 2002). All observational data have measurement errors. These are associated with instrumental errors, instrument siting, fetch, flux corrections, lack of energy balance closure, and neglected terms etc. (e.g., Offerle et al. 2005; Grimmond 2006). This dataset was chosen as it has been used previously by a number of groups to evaluate their models (Grimmond and Oke 2002; Masson et al. 2002; Best et al. 2006; Krayenhoff and Voogt 2007; Oleson et al. 2008a). This meant that parameter values were reasonably well known. Also, the observed fluxes were provided so no model/group had an advantage from previous knowledge of this data.

The observations used in the evaluation consist of Q^*, Q_H, and Q_E plus ΔQ_S determined as a residual (Grimmond and Oke 1999a). During the observations (14 days in August 1992), the area was in drought and there was an irrigation ban in the city that was adhered to (Grimmond and Oke 1999c). The area is characterized by little vegetation (<5% plan area cover) and the soil moisture was very low at the time of data collection (Grimmond and Oke 1999a,c), making Q_E at this site small relative to the other fluxes (Table 4). The summertime conditions are expected to be associated with low Q_F as the area did not have extensive use of air conditioning or other significant sources of Q_F. This would be expected to be more significant in the winter but is not considered here as no observational data were available.

The purpose of this comparison is not to identify the best model, but to understand model errors related to the type of approach taken (Figs. 1, 2). Each model was assigned a random identifier number, which is used in the subsequent analysis of the results to ensure anonymity. The returned simulation data from each of these models were used to perform a series of statistical analyses to evaluate model performance (Table 5).

5. Results from VL92

Using the VL92 dataset, 33 different models/versions of models were analyzed (Table 1). Modeling groups assigned parameter values and initial state conditions they thought appropriate. Of the 33 participants, 20 chose to rerun their models subsequent to their initial submission and based on developments of their models during the period of the model comparison, thereby improving performance. Of those who did, 16, three, and one groups reran their models once, twice, and three times, respectively, with a decrease in the root-mean-square error (RMSE) in all cases except for the minimum values for Q_E and ΔQ_S, which remain the same (Table 4). The remainder of this paper evaluates the performance based on the final run results only.

As noted, this site has been used to evaluate model performance in previous studies (Table 4). These evaluations are not directly comparable to the current data as the same forcing data were not used in all the studies, and the time periods are not consistent, unlike the current comparison where all models have followed an identical protocol. However, comparing those results with the "final" runs presented here we can see that the results are similar. As with the overall cohort of models participating in the International Urban Model Comparison, there is some suggestion that model performance may have improved in the current (final) runs.

For Q^* the models, on average, have a smaller systematic RMSE (RMSE_S) than unsystematic RMSE (RMSE_U; Table 4). However, the maximum RMSE_S (81.9 W m^{-2}) is the same order of magnitude as the maximum RMSE_U (80.7 W m^{-2}), suggesting there are problems that could be fixed, for example, by changing parameter values. For Q_H the mean and maximum RMSE_S are larger than the RMSE_U, also suggesting that model results might be improved.

The ranked performance of the individual models, based on RMSE calculated for the 312-h dataset, for the four fluxes is shown in Fig. 6. No individual model performs best or poorest for all fluxes. For each flux, when models are ordered from best to poorer performance, in the better performing models there are small differences in RMSE. However, there is a point of step drop in performance: for Q^* five models performing less well; for Q_H, 15 models show distinctly poorer performance.

The encouraging performance for Q_E, with small RMSE values and only two models performing noticeably poorer, is a function of its small flux (Table 4). When a normalized Taylor (2001) plot is considered (Figs. 6e–h), where the ideal model would fall at the square (the observations), Q_E is the least well modeled (Fig. 6h). For Q^*, the models cluster most closely to the observed value, except for the five outliers already identified. Again for Q^*, all models have a correlation coefficient (r) of greater than 0.95 except for one, which has an r value of over 0.9. It is interesting that there is less of a step drop in ΔQ_S model performance but an almost constant correlation coefficient for all models (~0.9). Also Q_H has an almost constant correlation coefficient

TABLE 4. Summary of the mean, maximum, and minimum statistical performance (see Table 5 for definitions of statistics) across 33 models when compared with the VL92 dataset for all hours (n = 312 h). Also, RMSE statistics are displayed for the first run of the models. RMSE values from previous evaluations using VL92 data for all hours (note these are not directly comparable as different time periods and forcing data are used in some cases): 1 indicates Grimmond and Oke (2002); 2 indicates Masson et al. (2002) for periods 223–236, 225–231, and 232–236), respectively (the asterisk denotes combined $Q_H + Q_E$); 3 indicates Best et al. (2006) tile 1 and 2, respectively; 4 indicates Krayenhoff and Voogt (2007) LI1 (original simulation) and LI2 (with parameter adjustments), respectively, for TUF3D for 0300–2100 on day 227; 5 indicates Oleson et al. (2008b) for periods 225–231 and 232–236, respectively.

Statistic		Q^*	Q_H	ΔQ_S	Q_E
$\bar{x}_{obs}$ (W m^{-2})		131.2	71.8	42.4	15.5
σ_{obs} (W m^{-2})		217.6	89.2	133.0	21.0
$\bar{x}_{mod}$ (W m^{-2})	Max	193.6	208.5	83.9	30.7
	Min	84.4	49.8	−15.3	0
	Mean	133.7	113.6	13.3	7.4
σ_{mod} (W m^{-2})	Max	268.6	197.4	187.3	34
	Min	153.8	67.7	41.5	0
	Mean	231	120.5	119.2	8.8
R^2	Max	0.99	0.85	0.88	0.39
	Min	0.85	0.61	0.45	0.01
	Mean	0.98	0.8	0.79	0.25
RMSE prior runs (W m^{-2})	1	—	49	—	20
	2	59, 57, 59	76, 97, 50*	91, 105, 66	—
	3	69, 71	56, 43	103, 81	27, 24
	4	40.2, 31.1	138.5, 107.4	109.9, 98.1	—
	5	34, 34	81, 49	86, 59	16, 23
*RMSE first run (W m^{-2})	Max	177.9	233.3	311.4	157.4
	Min	28.4	39.3	49.1	17.2
	Mean	58.4	95.5	87.8	30.3
RMSE final run (W m^{-2})	Max	92.3	183.1	115.7	31.5
	Min	22.1	36.8	49.1	17.2
	Mean	47	81.7	77.8	23
RMSE$_S$ (W m^{-2})	Max	81.9	163.8	111.5	26.1
	Min	4.2	6.8	16	4.8
	Mean	30.3	58.9	54.8	19.8
RMSE$_U$ (W m^{-2})	Max	80.7	81.8	79.3	27.6
	Min	18.1	32.2	15.7	0
	Mean	33.6	53	50.6	7.4
MAE (W m^{-2})	Max	76.6	136.7	89.7	21.4
	Min	15.4	24.7	33.1	11.5
	Mean	37	57.1	57.5	15.6
MBE (W m^{-2})	Max	62.4	136.7	41.4	15.2
	Min	−46.9	−22	−57.7	−15.6
	Mean	2.7	44.1	−30.9	−8.3
d	Max	1	0.95	0.96	0.78
	Min	0.94	0.66	0.58	0.42
	Mean	0.99	0.86	0.88	0.54

for all models (~0.9). Based on the index of agreement d, on average model performance is best for Q^*, followed by ΔQ_S, Q_H, and Q_E (Table 4). This ranking is retained when the best overall performance (maximum d) of any model for each flux is considered.

Models need to respond to changes in exchange processes through the course of the day. Of interest, for example, is whether they resolve peak radiant and turbulent heat fluxes during the day as well as fluxes at night when shortwave radiation does not need to be considered. When the data are analyzed by time of day, RMSE is larger during the day (Fig. 7) as expected because of its larger absolute magnitude. Figure 7 shows results for three time periods: 1) day (1 h after $Q^* \geq 0$ W m^{-2}), 2) night (1 h after $Q^* \leq 0$ W m^{-2}), and 3) transition (remaining hours when Q^* is going through 0 W m^{-2}). The five models with the largest RMSE for daytime Q^* (Fig. 7), are the same as those for all hours (Fig. 6), although the ranked order differs slightly. The transition hours are particularly problematic for these models. The two poorest performing in the daytime are among the six-poorest performing at night.

The observed fluxes of Q_H may be underestimated on some occasions because of advection caused by sea

TABLE 5. Statistics used to analyze model performance (Willmott 1981; Jacobson 1999): P_i and O_i = predicted and observed values; n = number of data points; $e_i = P_i - O_i \hat{P}_i = a + bO_i$ (where a and b are the intercept and slope of regression line between O and P).

Statistic		Description/equation
$\overline{x}$	Mean	$\overline{P} = \frac{\sum_{i=1}^{n} P}{n}$
σ	Std dev	$\sigma_p = \sqrt{\frac{\sum_{i=1}^{n}(P_i - \overline{P})^2}{n-1}}$
R	Correlation coef (Pearson's)	$R = \frac{\sum_{i=1}^{n} O_i P_i - \frac{\sum_{i=1}^{n} O_i \sum_{i=1}^{n} P_i}{n}}{\sqrt{\left[\sum_{i=1}^{n} O_i^2 - \frac{\left(\sum_{i=1}^{n} O_i\right)^2}{n}\right]\left[\sum_{i=1}^{n} P_i^2 - \frac{\left(\sum_{i=1}^{n} P_i\right)^2}{n}\right]}}$
R^2	Coef of determination	R^2
RMSE	Root-mean-square error	$\mathrm{RMSE} = \left[n^{-1}\sum_{i=1}^{n}(e_i)^2\right]^{0.5}$
RMSE_S	Systematic RMSE	$\mathrm{RMSE}_S = \left[n^{-1}\sum_{i=1}^{n}(\hat{P}_i - O_i)^2\right]^{0.5}$
RMSE_U	Unsystematic RMSE	$\mathrm{RMSE}_U = \left[n^{-1}\sum_{i=1}^{n}(\hat{P}_i - P_i)^2\right]^{0.5}$
MAE	Mean absolute error	$\mathrm{MAE} = n^{-1}\sum_{i=1}^{n}\lvert e_i\rvert$
MBE	Mean bias error	$\mathrm{MBE} = n^{-1}\sum_{i=1}^{n} e_i = \sum\overline{P} - \overline{O}$
d	Index of agreement	$d = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(\lvert O_i - \overline{O}\rvert + \lvert P_i - \overline{O}\rvert)^2}$

breezes (Masson et al. 2002). For Q_H the daytime errors are largest. At night the models generally do well almost across the board but the absolute values of the fluxes are smaller (Fig. 7). The daytime RMSE for Q_H is larger than for Q^* for all models. The $\mathrm{RMSE_S}$ tends to be greater for Q_H than for Q^*. For the most poorly performing models, $\mathrm{RMSE_s}$ is generally larger than $\mathrm{RMSE_U}$ (Fig. 7, circles plot above triangles).

Using the model classifications (Figs. 1, 2) we can evaluate whether particular approaches result in clear improvements in performance. It should be noted that the options used by groups were not always their most complex (cf. capability with VL92 options used in Figs. 2, 3). Two sets of statistics are used: RMSE and the mean bias error (MBE) for day and night (Figs. 8, 9) with results for each model shown as a point for each class and category (Fig. 2). The range, interquartile range (IQR), and mean and median performance of the category within the class can be compared. Perfect performance would have an RMSE and MBE of 0 W $\mathrm{m^{-2}}$. Given the relative magnitude of the MBE for nighttime Q_E ($<|12|$ W $\mathrm{m^{-2}}$), these results are not considered further here.

First, the method to represent vegetation (V class 1) is considered. Of the 18 models that have the ability to include vegetation as a separate tile (Vs; Fig. 2), five did not. Six additional models have integrated vegetation (Vi) within their urban surface. For the VL92 runs, a total of 14 models do not consider vegetation (Vn). The IQR of RMSE (bars on Fig. 8) is smaller in the daytime for Q^*, Q_H, and ΔQ_S when vegetation is included as a separate tile (Vs). In the daytime, not including vegetation (Vn) results in the largest RMSE medians (Q_H = 181, ΔQ_S = 136, Q^* = 59, and Q_E = 36 W $\mathrm{m^{-2}}$) and MBE medians (Q_H = 158 W $\mathrm{m^{-2}}$, $\Delta Q_S = -107$, $Q^* = 42$, and $Q_E = -28$). For daytime Q_H and Q^*, the tiled approach (Vs) has the smallest RMSE (median = 71 and 46 W $\mathrm{m^{-2}}$, respectively) and MBE (median = 18 and -14 W $\mathrm{m^{-2}}$, respectively), whereas the integrated vegetation (Vi) has the lowest individual RMSE values for Q^* and ΔQ_S. For daytime Q_E, the RMSE and MBE are best for Vi (median = 27 and 3 W $\mathrm{m^{-2}}$, respectively). At night for Q_H, the performance is poorest for those models that assume a separate tile (Vs; median RMSE = 19 W $\mathrm{m^{-2}}$, MBE = 17 W $\mathrm{m^{-2}}$) and best for Vi models

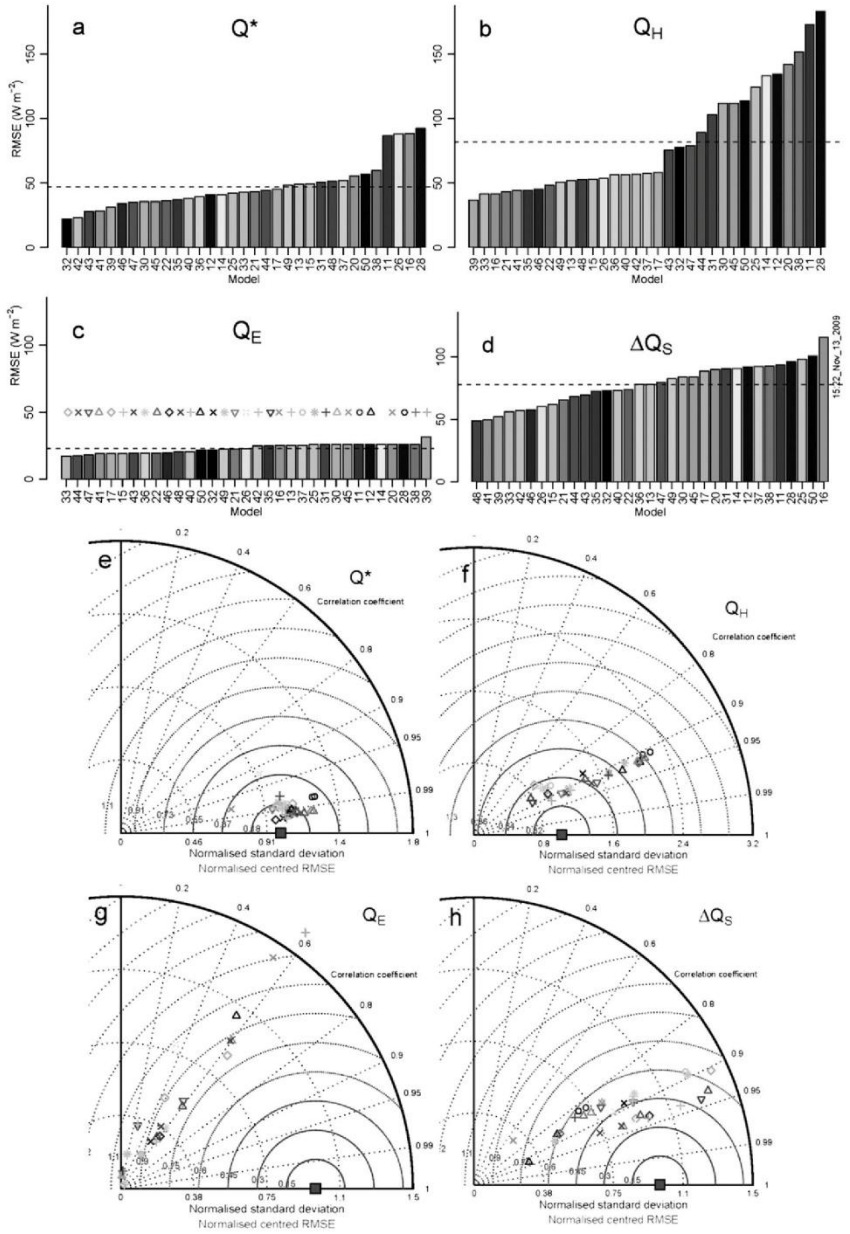

FIG. 6. (top) Ranked RMSE (W m^{-2}) and (bottom) normalized Taylor diagrams associated with each model for the whole time period. Each model is randomly assigned a number and symbol. The key for the symbols is shown in (c). (a),(e) Net all-wave radiation, (b),(f) turbulent sensible heat, (c),(g) latent heat, and (d),(h) storage heat fluxes. The dotted line is the mean RMSE. The Taylor plots display the correlation coefficient in relation to the polar axis comparing hourly values, the normalized standard deviation in relation to the horizontal axis and the normalized RMSE in relation to the internal circular axes (Taylor 2001); (N = 312 h).

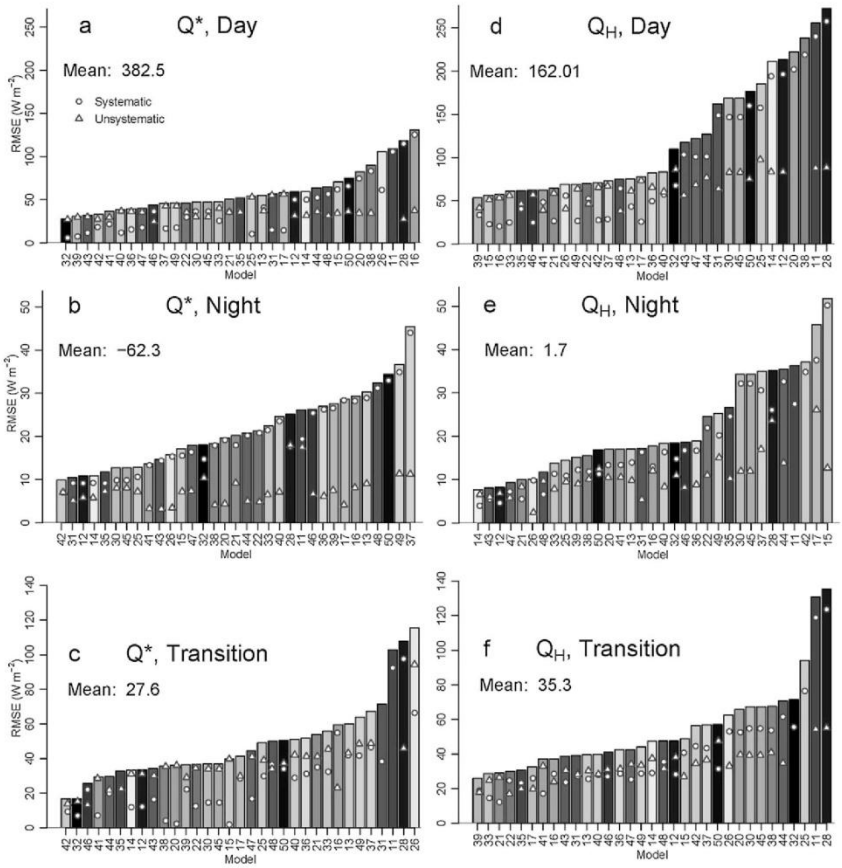

FIG. 7. Ranked RMSE (W m^{-2}) for (left) net all-wave radiation and (right) turbulent sensible heat flux by time of day (see text) for (a),(d) day, (b),(e) night, and (c),(f) transition time periods. Circles and triangles are $RMSE_S$ and $RMSE_U$, respectively. The mean observed flux (W m^{-2}) is for each period is given.

(median RMSE = 14 W m^{-2}, MBE = −2 W m^{-2}). However, for Q^*, Vi and Vn have similar performance (median RMSE Vi = 16, Vn = 19; MBE: Vi = −11, Vn = −8 W m^{-2}).

Examining the combination of model characteristics (Fig. 3) shows that for those that do not take into account vegetation, Vn, share only one common characteristic: their calculation of ΔQ_S via conduction or net radiation (class 8, Sc, Sn). However, many models that do include vegetation (Vs) also use this approach to heat conduction (Sc), so this is not likely to be a primary co-explanation. Not including vegetation even in this area where there is very little, and where the measured Q_E is small relative to the other fluxes, appears to impact the ability to model Q^* and Q_H, with a resulting poor performance also in ΔQ_S.

The VL92 site also has low Q_F. Most groups assumed it is negligible (ANn) with only seven groups explicitly including the flux (Fig. 2). Those that have considered it have taken a wide range of approaches but because of the small numbers they are grouped together into one class for analysis (ANm). Similarly, different temporal approaches to modeling Q_F (Tf, Tv) are used but the small number of models per class means analysis is the same and so is not shown separately. In the daytime, median RMSE and MBE are smallest for all fluxes when Q_F is ignored (ANn). This differs for nighttime fluxes however, where ANm models have

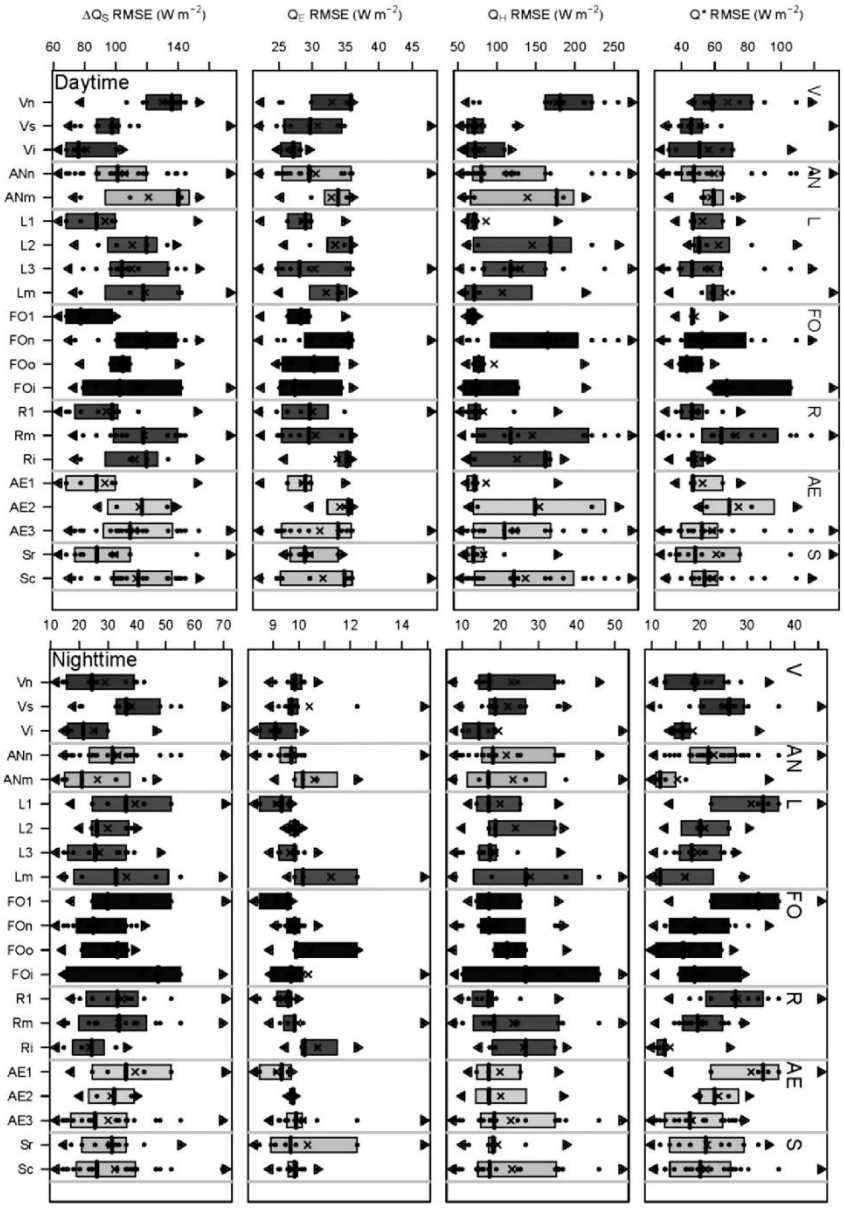

FIG. 8. RMSE (W m^{-2}) for each of the seven categories (see Fig. 2 for key) (left to right) ΔQ_S, Q_E, Q_H, and Q^* for (top) day and (bottom) night. Each dot is a model, the shaded bar shows the 25th and 75th percentile, and the line indicates the median and the "x" the mean. The maximum and minimum are indicated by the triangles. Note that the vertical scale varies between graphs. Mean observed fluxes for Q_E is 28.6 (day) and 3.9 W m^{-2} (night) and for ΔQ_S is 191.7 (day) and −70.7 W m^{-2} (night). For Q^* and Q_H see Fig. 7.

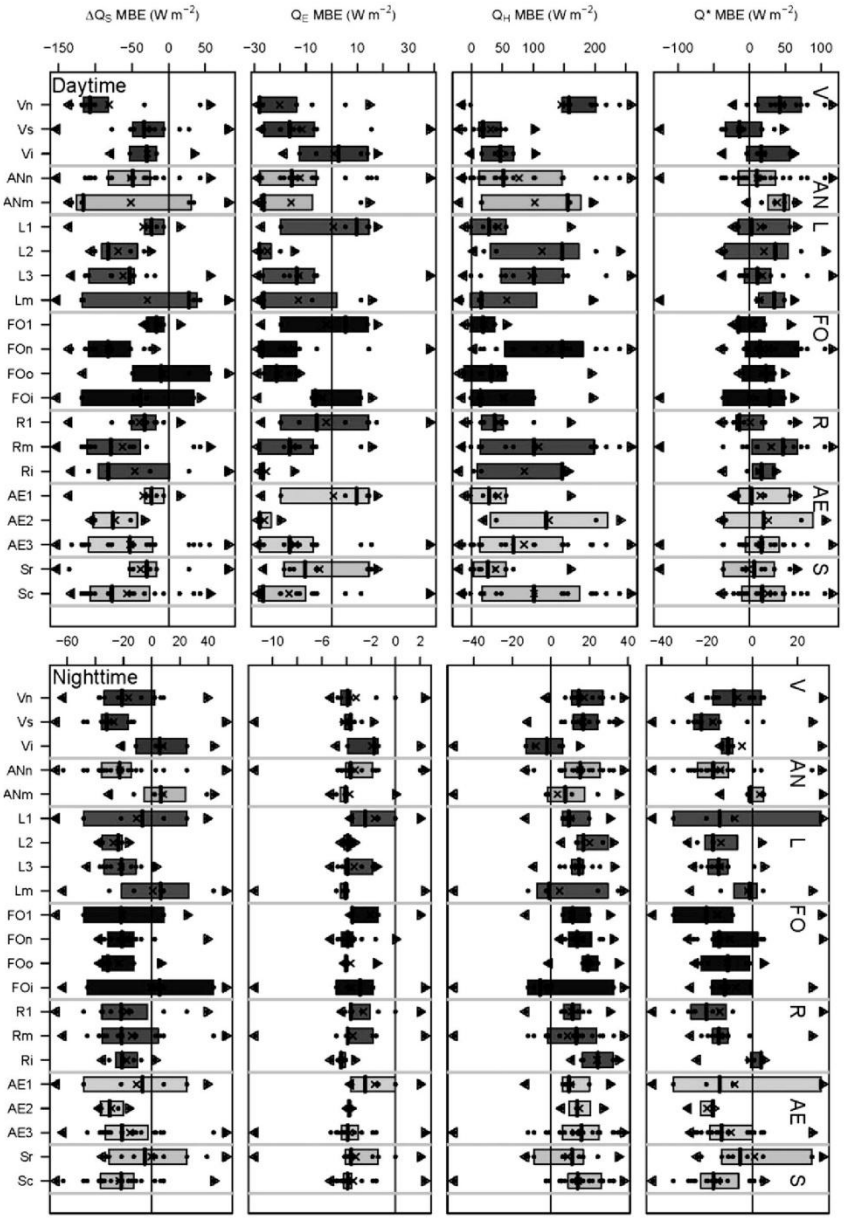

FIG. 9. As in Fig. 8 but for the MBE (W m^{-2}).

the smallest RMSE and absolute MBE for all fluxes except Q_E.

The model combinations in Fig. 3 show that those models that use an internal temperature (ANi) tend to have a fixed or variable temporal variation in Q_F (class 3, Tf, Tv), an urban morphology that is multilayered (L4–L7), and a surface albedo–emissivity that has three or more facets (class 7, AEf).

The urban morphology (class 4, L) has a relatively large within-class difference (range of median RMSE: 98 W m^{-2} and MBE: 130 W m^{-2}) for the daytime Q_H. For both RMSE and MBE, there is no clear best performer of models across all fluxes (best for RMSE median: Q^* = L3, Q_H = Lm and L1, Q_E = L3, ΔQ_S = L1; best for MBE median: ΔQ_S = L1, Q_H = Lm, Q^* = L1, Q_E = L1; Figs. 8, 9). At night, multilayer models (Lm = L4–L7) perform best for Q^*, Q_H, and ΔQ_S based on MBE (median $Q^* = -1$, $Q_H = -1$, ΔQ_S = 6 W m^{-2}). The urban morphology classes have few common characteristics, although all L1 models use a single reflection and a bulk albedo and emissivity. In addition, and by definition, L3 and all Lm models have three facets for albedo and emissivity.

With respect to the categorization based on facets and orientation (FO class 5), the largest difference is for the simulation of daytime Q_H (difference between category medians ΔRMSE of 96 W m^{-2}, ΔMBE = 129 W m^{-2}). Those that treat the surface as a whole (FO1) have the lowest daytime RMSE for Q_H and ΔQ_S (although for Q_H, median RMSE for FO1, FOo, and FOi differ by <8 W m^{-2} while it is lowest for FOo and Q^* and for FOi and Q_E). At night, the lowest median RMSE is: Q^* = FOo, Q_H = FO1, and FOn, Q_E = all groups equal, ΔQ_S = FOn. There is no consistency in groupings with the smallest daytime MBEs (Q^* = FOn, Q_H = FOi, Q_E = FO1, ΔQ_S = FOo). Except for Q^*, during the daytime, models that simulate a canyon but have no associated orientation (FOn), have the largest biases (Q_H: positive bias, Q_E and ΔQ_S: negative bias) and these are likely to be complementary. At night, models that incorporate orientation and intersections (FOi) have the smallest bias, again except for Q^*, where it is FOo models (although differing by just 1 W m^{-2} when compared with FOi). In the daytime, for Q_H, the median RMSE performance improves from FOn, FOo, FOi, and FO1 (165, 77, 74, and 69 W m^{-2}, respectively); and for Q^*, improves from FOi, FOn, FO1, and FOo (67, 52, 46, and 43 W m^{-2}). The unique combinations that these categories of models have in common include those that treat the surface as a whole (FO1), have no anthropogenic heat fluxes calculated (ANn) and, obviously, have just a slab surface morphology, single reflections, and a bulk albedo and emissivity. Models that include orientation (FOo, FOi) all assume three or more facets for albedo and emissivity (AEf; as would be expected). Models without orientation (FOn) largely utilize conduction methods to calculate the storage heat flux (Sc).

When the models are classified based on the number of reflections used, there are large within class differences (ΔRMSE = 89 W m^{-2} for daytime Q_H; Fig. 8). This is also the largest difference for the MBE (ΔMBE 109 W m^{-2}; Fig. 9). During the day, models with a single reflection scheme (class 6, R1) perform best for all fluxes except Q_E (median RMSE ΔQ_S = 98, Q_H = 73, Q^* = 46 W m^{-2}). The daytime MBE is smallest for Q^* models that calculate single reflections (Rs; median MBE $Q^* = -14$). Generally, during the daytime the models that have infinite reflections (Ri) perform least well for Q_H and Q_E, (median MBE Q_H = 147, $Q_E = -27$ W m^{-2}; median RMSE Q_H = 162, Q_E = 35 W m^{-2}); there are also negative median MBEs for all classes for Q_E and ΔQ_s, while Q_H and Q^* have a positive bias, with the exception of the single reflection class and Q^*. This suggests that the single reflection models may not allow enough radiation to be absorbed in comparison with observations. For ΔQ_S and Q_H, RMSE increases with the number of reflections modeled.

At night, models using increasing numbers of reflections have smaller RMSE for Q^* (Q^*: Ri = 13, Rm = 20, R1 = 28 W m^{-2}); whereas, the situation reverses for Q_H and Q_E, with those modeling fewer reflections yielding better results (Q_H: Ri = 27, Rm = 18, R1 = 17 W m^{-2}). For the calculation of Q^* at night, the Ri type models perform best with the lowest median RMSE and MBE (RMSE = 13, MBE = 4 W m^{-2}). However, as for daytime, superior performance for one flux is accompanied by poorer performance in another. All approaches have a similar sized negative MBE for nocturnal ΔQ_S (median from −21 to −22 W m^{-2}). The MBE for single reflections suggests that the surface temperature is too high, but correcting the bias during the daytime is likely to increase the nocturnal surface temperature, so there may be other issues with the models that use this method. Compensation also occurs between Q^* and Q_H most particularly at night. All schemes with infinite reflections (Ri) have three facets for albedo and emissivity (AEf).

The differences within groups of models are amongst the greatest when stratified based on specification of albedo/emissivity (class 7, AE). In general, using a bulk albedo–emissivity (AE1) results in better performance for all fluxes during the day based on median RMSE and MBE (median MBE $\Delta Q_S = -23$, $Q^* = 3$, $Q_E = 10$, Q_H = 28 W m^{-2}). Models using two facets (AE2) tend to have the poorest daytime performance (except for Q^* where median MBE for all groups is similar). At night, the differences in median MBE are smaller (Q_H:

AE2 = 14, AEf = 16, AE1 = 9; Q^*: AE1 = −14, AEf = −13, AE2 = −17 W m^{-2}). In this evaluation, where buildings are small and widely spaced, the ability to distinguish different facet characteristics of albedo and emissivity is not important. However, where buildings are taller, more tightly spaced, and/or constructed with very different materials, this result may not necessarily be the same. It is also important to remember that depending on the intended application, the ability to change facet material characteristics may be very important; for example, for scenario testing (e.g., for urban heat island mitigation).

Classifying models based on method used to calculate ΔQ_S (S class 8) has a relatively small difference in the median RMSE and MBE for all fluxes. Again the biggest difference in performance is associated with daytime Q_H (52 and 74 W m^{-2} for RMSE and MBE). The daytime Q^* differences are 6 and 12 W m^{-2} for RMSE and MBE, respectively; these are the smallest within-group differences in median for Q^* across the classes. The residual method (Sr) performs better during the daytime for all fluxes except Q^* daytime (median MBE: Q_H = 27, Q^* = 6, Q_E = −11, ΔQ_S = −30 W m^{-2}) and for all nighttime fluxes (median MBE: Q_H = 11, Q_E = 4, ΔQ_S = −5, Q^* = −5 W m^{-2}). The Sc models often assume three facets (AEf) without orientation (FOn).

If the 31 different classes are considered, the best performance during the daytime for Q^* is from the FOo class (median RMSE of 43 W m^{-2}). There are two classes with an absolute median MBE of ≤3 W m^{-2} (L1, AE1). There are six models with both these characteristics (Fig. 3). For daytime Q_H, there are four classes with an MBE of <20 W m^{-2} (Vs, Lm, FO1, FOi). There is only one model with all of these (viz., Vs, Lm, FOi). The best overall performance for daytime Q_H, based on median RMSE, has a value of 69 W m^{-2} (FO1), but there are seven other classes within 4 W m^{-2} of this (Vs, Vi, Ls, Lm, R1, AE1, Sr) and three additional classes within 7 W m^{-2} (ANm, FOo, FOi), thereby accounting for all seven major classes (Fig. 2). No models have all of these characteristics, while two have five of them but do not generally fall within the group of best-performing models.

At night, best performance for Q^* is associated with ANm, Lm, Ri (median RMSE 11–13 W m^{-2} and/or median MBE < |4| W m^{-2}) and for Q_H with Vi, Lm, and FOi (median MBE = −2, −1, and −6, median RMSE = 14, 27, and 27 W m^{-2}, respectively). The Sr and Sc models have a similarly good RMSE (17–18 W m^{-2}).

For daytime Q_E, best overall performance is from Ls, FO1, AE1, Vi, FOi, and R1 (median MBE < |10| W m^{-2}). For daytime ΔQ_S, models with median RMSE < 96 W m^{-2} are Vi, Ls, FO1, AE1, and Sr but based on the absolute MBE, the best-performing models are FOo (median MBE = ≤|10| W m^{-2}), and FO1, AE1, Sr, Vi, Lm, and Ls (median MBE < |30| W m^{-2}). At night, Lm, ANm, FOi, Sr, and Vi models perform well based on median MBE and RMSE (<|4| and/or <22 W m^{-2}).

6. Conclusions

Urban surface–atmospheric exchanges are modeled for a wide variety of applications. The large set of models, examined here, have a range of approaches, complexities, and parameter requirements. Through the first stage of the first international model comparison reported here, significant model developments have taken place and improvements in model performance have resulted.

Evaluation of 33 models, with Vancouver (VL92) data, shows that generally models have best overall capability to model Q^* and least capability to model Q_E (order Q^*, ΔQ_S, Q_H, and Q_E; Table 4). No model performs best or worst for all fluxes. In particular, it seems to be difficult to minimize both Q^* and Q_H errors. There is evidence that some classes of models perform better for individual fluxes but not overall. Typically, those that perform best during daytime do not perform best at night.

The daytime RMSE for Q_H is larger than for Q^* for all but four models. These four are characterized as having amongst the four largest Q^* RMSE values. For RMSE$_S$, there is the tendency for Q_H errors to be greater than for Q^*, although there are more cases where the errors are similar. The unsystematic errors are generally smaller than systematic errors, particularly for the most poorly performing models. For most models, Q_H has a positive MBE, which observational errors may contribute to.

Seven characteristics (relating to vegetation, Q_F, morphology, facets and orientations, reflection, albedo and emissivity, and ΔQ_S) are used to classify each model. Some of the greatest differences in model performance are found between classes of model that treat vegetation and reflections differently. Some of the smallest differences relate to approaches used to calculate the ΔQ_S followed by urban morphology. Not including vegetation, even at a site with limited vegetation, yields the poorest performance for all fluxes during the day (in terms of RMSE) and for Q_E at night. During the day, median RMSE for models that do not include Q_F is similar (or better) than for those that do. However, at night, median RMSE for models, which include Q_F shows better performance for Q^*, Q_H, and ΔQ_S. Models that account for urban morphology orientation, and also intersections, often have slightly better performance than schemes that do not (e.g., Q_H in the daytime). The addition of intersections, however, does not always improve performance appreciably and in some cases has a negative impact on model performance.

The results for reflection schemes vary between day and night and with statistical measure (RMSE or MBE). In general, using a bulk albedo–emissivity results in better performance for all fluxes during the day. Classifying based on method used to calculate ΔQ_S has the smallest difference in the median of the RMSE and MBE of all classes. The residual method performs better during the day for all fluxes, while at night, differences are less significant. Class combinations show no models display all characteristics associated with strongest performance, although two display a large proportion of these. In general, the simpler models perform as well as the more complex models based on all statistical measures.

These results are based on a short time series for one urban location. In phase 2, the same models will be evaluated using a second dataset (Grimmond et al. 2009b). These results raise a number of questions that will be considered, with different flux partitioning, a wider range of conditions, and a longer time series. Of particular interest is whether the same models and classes perform well; whether the relative ability to model the individual fluxes remain the same; and whether it is possible for any class of models to minimize errors in both Q^* and Q_H.

Acknowledgments. Funded (Grimmond) by the Met Office (P001550), European Union Framework 7 (7 FP7-ENV-2007-1): MEGAPOLI (212520), BRIDGE (211345). This is a contribution to COST728 (Enhancing mesoscale meteorological modeling capabilities for air pollution and dispersion applications). We thank all involved in the original dataset collection (especially James Voogt and Tim Oke) and their funding agencies. Funding from BSIK-COM29 (Steeneveld), CATER 2006-2202 (Baik), and other agencies that support all the time contributed by participating groups is acknowledged. Color versions of the figures can be obtained by e-mailing corresponding author Sue Grimmond.

REFERENCES

Best, M. J., 1998: Representing urban areas in numerical weather prediction models. *Proc. Second Urban Environment Symp.,* Albuquerque, NM, Amer. Meteor. Soc., paper 10A.2.

——, 2005: Representing urban areas within operational numerical weather prediction models. *Bound.-Layer Meteor.,* **114,** 91–109.

——, C. S. B. Grimmond, and M. G. Villani, 2006: Evaluation of the urban tile in MOSES using surface energy balance observations. *Bound.-Layer Meteor.,* **118,** 503–525.

Chen, F., and J. Dudhia, 2001: Coupling an advanced land–surface/hydrology model with the Penn State/NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Mon. Wea. Rev.,* **129,** 569–585.

——, H. Kusaka, M. Tewari, J. Bao, and H. Hirakuchi, 2004: Utilizing the coupled WRF/LSM/Urban modeling system with detailed urban classification to simulate the urban heat island phenomena over the Greater Houston area. *Proc. Fifth Conference on Urban Environment,* Vancouver, BC, Canada, Amer. Meteor. Soc., paper 9.11.

Ching, J., and Coauthors, 2009: National Urban Database and Access Portal Tool. *Bull. Amer. Meteor. Soc.,* **90,** 1157–1168.

Clarke, J. A., 1985: *Energy Simulation in Building Design.* Adam Hilger, 362 pp.

Dandou, A., M. Tombrou, E. Akylas, N. Soulakellis, and E. Bossioli, 2005: Development and evaluation of an urban parameterization scheme in the Penn State/NCAR Mesoscale model (MM5). *J. Geophys. Res.,* **110,** D10102, doi:10.1029/2004JD005192.

Dupont, S., and P. G. Mestayer, 2006: Parameterization of the urban energy budget with the Submesoscale Soil Model. *J. Appl. Meteor. Climatol.,* **45,** 1744–1765.

——, ——, E. Guilloteau, E. Berthier, and H. Andrieu, 2006: Parameterization of the urban water budget with the Submesoscale Soil Model. *J. Appl. Meteor. Climatol.,* **45,** 624–648.

Essery, R. L. H., and D. B. Clark, 2003: Developments in the MOSES 2 land-surface model for PILPS 2e. *Global Planet. Change,* **38,** 161–164.

——, M. J. Best, R. A. Betts, P. M. Cox, and C. M. Taylor, 2003: Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *J. Hydrometeor.,* **4,** 530–543.

Feddema, J. J., K. W. Oleson, G. B. Bonan, L. O. Mearns, L. E. Buja, G. A. Meehl, and W. M. Washington, 2005: The importance of land-cover change in simulating future climates. *Science,* **310,** 1674–1678.

Fortuniak, K., B. Offerle, and C. S. B. Grimmond, 2004: Slab surface energy balance scheme and its application to parameterisation of the energy fluxes on urban areas. *NATO ASI,* Kiev, Ukraine. [Available online at www.met.rdg.ac.uk/urb_met/NATO_ASI/talks.html.]

——, ——, and ——, 2005: Application of a slab surface energy balance model to determine surface parameters for urban areas, Lund eRep. *Phys. Geogr.,* **5,** 90–91.

Grimmond, C. S. B., 1992: The suburban energy balance: methodological considerations and results for a mid-latitude west coast city under winter and spring conditions. *Int. J. Climatol.,* **12,** 481–497.

——, 2006: Progress in measuring and observing the urban atmosphere. *Theor. Appl. Climatol.,* **84,** 3–22.

——, and T. R. Oke, 1986: Urban water balance II: Results from a suburb of Vancouver, B.C. *Water Resour. Res.,* **22,** 1404–1412.

——, and ——, 1991: An evaporation–interception model for urban areas. *Water Resour. Res.,* **27,** 1739–1755.

——, and ——, 1999a: Heat storage in urban areas: Observations and evaluation of a simple model. *J. Appl. Meteor.,* **38,** 922–940.

——, and ——, 1999b: Aerodynamic properties of urban areas derived from analysis of surface form. *J. Appl. Meteor.,* **38,** 1262–1292.

——, and ——, 1999c: Rates of evaporation in urban areas. Impacts of urban growth on surface and ground waters. International Association of Hydrological Sciences Publication 259, 235–243.

——, and ——, 2002: Turbulent heat fluxes in urban areas: Observations and local-scale urban meteorological parameterization scheme (LUMPS). *J. Appl. Meteor.,* **41,** 792–810.

——, ——, and D. G. Steyn, 1986: Urban water balance I: A model for daily totals. *Water Resour. Res.,* **22,** 1404–1412.

——, H. A. Cleugh, and T. R. Oke, 1991: An objective urban heat storage model and its comparison with other schemes. *Atmos. Environ.,* **25B,** 311–326.

——, C. Souch, and M. Hubble, 1996: The influence of tree cover on summertime energy balance fluxes, San Gabriel Valley, Los Angeles. *Climate Res.,* **6,** 45–57.

——, J. A. Salmond, T. R. Oke, B. Offerle, and A. Lemonsu, 2004: Flux and turbulence measurements at a dense urban site in Marseille: Heat, mass (water, carbon dioxide) and momentum. *J. Geophys. Res.,* **109,** D24101, doi:10.1029/2004JD004936.

——, and Coauthors, 2009a: Urban Surface Energy Balance Models: Model characteristics and methodology for a comparison study. *Meteorological and Air Quality Models for Urban Areas,* A. Baklanov et al., Eds., Springer-Verlag, 97–124.

——, and Coauthors, 2009b: The International Urban Energy Balance Comparison Project: Initial results from phase 2. *Proc. Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, International Association for Urban Climate paper A11-6, 4 pp. [Available online at http://www.ide.titech.ac.jp/~icuc7/extended_abstracts/pdf/372811-1-090524190750-003.pdf.]

Guilloteau, E., 1998: Optimized computation of transfer coefficients in surface layer with different momentum and heat roughness length. *Bound.-Layer Meteor.,* **87,** 147–160.

Hamdi, R., and G. Schayes, 2007: Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities. *Atmos. Chem. Phys.,* **7,** 4513–4526.

——, and V. Masson, 2008: Inclusion of a drag approach in the Town Energy Balance (TEB) scheme: Offline 1D evaluation in a street canyon. *J. Appl. Meteor. Climatol.,* **47,** 2627–2644.

Hanna, S. R., and J. C. Chang, 1992: Boundary layer parameterizations for applied dispersion modeling over urban areas. *Bound.-Layer Meteor.,* **58,** 229–259.

——, and ——, 1993: Hybrid Plume Dispersion Model (HPDM) improvements and testing at three field sites. *Atmos. Environ.,* **27A,** 1491–1508.

Harman, I. N., and S. E. Belcher, 2006: The surface energy balance and boundary layer over urban street canyons. *Quart. J. Roy. Meteor. Soc.,* **132,** 2749–2768.

——, M. J. Best, and S. E. Belcher, 2004a: Radiative exchange in an urban street canyon. *Bound.-Layer Meteor.,* **110,** 301–316.

——, J. F. Barlow, and S. E. Belcher, 2004b: Scalar fluxes from urban street canyons. Part II: Model. *Bound.-Layer Meteor.,* **113,** 387–410.

Heiple, S., and D. J. Sailor, 2008: Using building energy simulation and geospatial modelling techniques to determine high resolution building sector energy consumption profiles. *Energy Build.,* **40,** 1426–1436.

Henderson-Sellers, A., and R. E. Dickinson, 1992: Intercomparison of land-surface parameterizations launched. *Eos, Trans. Amer. Geophys. Union,* **73,** 195–196.

——, Z. L. Yang, and R. E. Dickinson, 1993: The Project for Intercomparison of Land-surface Parameterization Schemes. *Bull. Amer. Meteor. Soc.,* **74,** 1335–1349.

——, P. Irannejad, K. McGuffie, and A. Pitman, 2003: Predicting land-surface climates—better skill or moving targets? *Geophys. Res. Lett.,* **30,** 1777, doi:10.1029/2003GL017387.

Ichinose, T., K. Shimodozono, and K. Hanaki, 1999: Impact of anthropogenic heat on urban climate in Tokyo. *Atmos. Environ.,* **33,** 3897–3909.

Irranejad, P., A. Henderson-Sellers, and S. Sharmeen, 2003: Importance of land-surface parameterization for latent heat simulation in global atmospheric models. *Geophys. Res. Lett.,* **30,** 1904, doi:10.1029/2003GL018044.

Jacobson, M. Z., 1999: *Fundamentals of Atmospheric Modeling.* Cambridge University Press, 656 pp.

Kanda, M., T. Kawai, M. Kanega, R. Moriwaki, K. Narita, and A. Hagishima, 2005a: A simple energy balance model for regular building arrays. *Bound.-Layer Meteor.,* **116,** 423–443.

——, ——, and K. Nakagawa, 2005b: A simple theoretical radiation scheme for regular building arrays. *Bound.-Layer Meteor.,* **114,** 71–90.

——, M. Kanega, T. Kawai, H. Sugawara, and R. Moriwaki, 2007: Roughness lengths for momentum and heat derived from outdoor urban scale models. *J. Appl. Meteor. Climatol.,* **46,** 1067–1079.

Kawai, T., M. Kanda, K. Narita, and A. Hagishima, 2007: Validation of a numerical model for urban energy-exchange using outdoor scale-model measurements. *Int. J. Climatol.,* **27,** 1931–1942.

——, M. K. Ridwan, and M. Kanda, 2009: Evaluation of the Simple Urban Energy Balance Model using selected data from 1-yr flux observations at two cities. *J. Appl. Meteor. Climatol.,* **48,** 693–715.

Kawamoto, Y., and R. Ooka, 2006: Analysis of the radiation field at pedestrian level using a meso-scale meteorological model incorporating the urban canopy model. Preprints, *Sixth Int. Conf. on Urban Climate,* Göteburg, Sweden, WMO, 446–449.

——, and ——, 2009a: Accuracy validation of urban climate analysis model using MM5 incorporating a multi-layer urban canopy model. *Extended Abstracts, Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, paper A10-1.

——, and ——, 2009b: Development of urban climate analysis model using MM5 Part 2 - Incorporating an urban canopy model to represent the effect of buildings (in Japanese). *J. Environ. Eng.,* **74,** 1009–1018.

Kikegawa, Y., Y. Genchi, H. Yoshikado, and H. Kondo, 2003: Development of a numerical simulation system toward comprehensive assessments of urban warming countermeasures including their impacts upon the urban buildings' energy-demands. *Appl. Energy,* **76,** 449–466.

Kondo, H., and F. H. Liu, 1998: A study on the urban thermal environment obtained through a one-dimensional urban canopy model (in Japanese). *J. Japan Soc. Atmos. Environ.,* **33,** 179–192.

——, Y. Genchi, Y. Kikegawa, Y. Ohashi, H. Yoshikado, and H. Komiyama, 2005: Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: Structure of the urban canopy model and its basic performance. *Bound.-Layer Meteor.,* **116,** 395–421.

Krayenhoff, E. S., and J. A. Voogt, 2007: A microscale three-dimensional urban energy balance model for studying surface temperatures. *Bound.-Layer Meteor.,* **123,** 433–461.

Kusaka, H., H. Kondo, Y. Kikegawa, and F. Kimura, 2001: A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Bound.-Layer Meteor.,* **101,** 329–358.

Lee, S.-H., and S.-U. Park, 2008: A vegetated urban canopy model for meteorological and environmental modelling. *Bound.-Layer Meteor.,* **126,** 73–102.

——, C.-K. Song, J.-J. Baik, and S.-U. Park, 2009: Estimation of anthropogenic heat emission in the Gyeong-In region of Korea. *Theor. Appl. Climatol.,* **96,** 291–303.

Lemonsu, A., C. S. B. Grimmond, and V. Masson, 2004: Modelling the surface energy balance of an old Mediterranean city core. *J. Appl. Meteor.,* **43,** 312–327.

——, and Coauthors, 2008: Overview and first results of the Montreal Urban Snow Experiment 2005. *J. Appl. Meteor. Climatol.,* **47,** 59–75.

Liu, Y., F. Chen, T. Warner, and J. Basara, 2006: Verification of a mesoscale data-assimilation and forecasting system for the Oklahoma City area during the Joint Urban 2003 Field Project. *J. Appl. Meteor. Climatol.,* **45,** 912–929.

Loridan, T., and Coauthors, 2010: Trade-offs and responsiveness of the single-layer urban canopy parametrization in WRF: An offline evaluation using the MOSCEM optimization algorithm and field observations. *Quart. J. Roy. Meteor. Soc.*, **136,** 997–1019.

Martilli, A., A. Clappier, and M. W. Rotach, 2002: An urban surface exchange parameterisation for mesoscale models. *Bound.-Layer Meteor.,* **104,** 261–304.

——, Y. A. Roulet, M. Junier, F. Kirchner, M. W. Rotach, and A. Clappier, 2003: On the impact of urban surface exchange parameterisations on air quality simulations: The Athens case. *Atmos. Environ.,* **37,** 4217–4231.

Masson, V., 2000: A physically-based scheme for the urban energy budget in atmospheric models. *Bound.-Layer Meteor.,* **41,** 1011–1026.

——, C. S. B. Grimmond, and T. R. Oke, 2002: Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities. *J. Appl. Meteor.,* **41,** 1011–1026.

Mitchell, V. G., R. G. Mein, and T. A. McMahon, 2001: Modelling the urban water cycle. *Environ. Model. Software,* **16,** 615–629.

Moriwaki, R., M. Kanda, H. Senoo, A. Hagishima, and T. Kinouchi, 2008: Anthropogenic water vapor emissions in Tokyo. *Water Resour. Res.,* **44,** W11424, doi:10.1029/2007WR006624.

Noilhan, J., and J. F. Mahfouf, 1996: The ISBA land surface parameterisation scheme. *Global Planet. Change,* **13,** 145–159.

Offerle, B., C. S. B. Grimmond, and T. R. Oke, 2003: Parameterization of net all-wave radiation for urban areas. *J. Appl. Meteor.,* **42,** 1157–1173.

——, ——, and K. Fortuniak, 2005: Heat storage and anthropogenic heat flux in relation to the energy balance of a central European city center. *Int. J. Climatol.,* **25,** 1405–1419.

Oke, T. R., 1988: The urban energy balance. *Prog. Phys. Geogr.,* **12,** 471–508.

——, R. Spronken-Smith, E. Jauregui, and C. S. B. Grimmond, 1999: The energy balance of central Mexico City during the dry season. *Atmos. Environ.,* **33,** 3919–3930.

Oleson, K. W., G. B. Bonan, J. Feddema, M. Vertenstein, and C. S. B. Grimmond, 2008a: An urban parameterization for a global climate model. Part I: Formulation and evaluation for two cities. *J. Appl. Meteor. Climatol.,* **47,** 1038–1060.

——, ——, ——, and ——, 2008b: An urban parameterization for a global climate model. Part II: Sensitivity to input parameters and the simulated heat island in offline simulations. *J. Appl. Meteor. Climatol.,* **47,** 1061–1076.

Pigeon, G., D. Legain, P. Durand, and V. Masson, 2007: Anthropogenic heat release in an old European agglomeration (Toulouse, France). *Int. J. Climatol.,* **27,** 1969–1981.

Pitman, A. J., A. Henderson-Sellers, and Z.-L. Yang, 1990: Sensitivity of regional climates to localized precipitation in global models. *Nature,* **346,** 734–737.

Porson, A., I. N. Harman, S. I. Bohnenstengel, and S. E. Belcher, 2009: How many facets are needed to represent the surface energy balance of an urban area? *Bound.-Layer Meteor.,* **132,** 107–128.

Roth, M., 2000: Review of atmospheric turbulence over cities. *Quart. J. Roy. Meteor. Soc.,* **126,** 941–990.

Rowley, F. B., A. B. Algren, and J. L. Blackshaw, 1930: Surface conductances as affected by air velocity, temperature and character of surface. *ASHRAE Trans.,* **36,** 429–446.

Ryu, Y.-H., J.-J. Baik, and S.-H. Lee, 2009: A new single-layer urban canopy model for use in mesoscale atmospheric models. *Proc. Seventh Int. Conf. on Urban Climate,* Yokohama, Japan, International Association for Urban Climate, paper A16-1. [Available online at http://www.ide.titech.ac.jp/~icuc7/extended_abstracts/pdf/374405-1-081210150038-004%28rev%29.pdf.]

Sailor, D. J., and L. Lu, 2004: A top-down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmos. Environ.,* **38,** 2737–2748.

Salamanca, F., and A. Martilli, 2009: A new Building Energy Model coupled with an Urban Canopy Parameterization for urban climate simulations—part II. Validation with one dimension off-line simulations. *Theor. Appl. Climatol.,* doi:10.1007/s00704-009-0143-8.

——, A. Krpo, A. Martilli, and A. Clappier, 2009: A new building energy model coupled with an urban canopy parameterization for urban climate simulations—part I. formulation, verification, and sensitivity analysis of the model. *Theor. Appl. Climatol.,* doi:10.1007/s00704-009-0142-9.

Shashua-Bar, L., and M. E. Hoffman, 2002: The green CTTC model for predicting the air temperature in small urban wooded sites. *Build. Environ.,* **37,** 1279–1288.

——, and ——, 2004: Quantitative evaluation of passive cooling of the UCL microclimate in hot regions in summer. *Build. Environ.,* **39,** 1087–1099.

Taylor, K., 2001: Summarizing multiple aspects of model performance in single diagram. *J. Geophys. Res.,* **106** (D7), 7183–7192.

Tremback, C. J., and R. Kessler, 1985: A surface temperature and moisture parameterization for use in mesoscale numerical models. *Proc. Seventh Conf. on Numerical Weather Prediction,* Montreal, Quebec, Amer. Meteor. Soc., 433–434.

UN, cited 2009: World urbanization prospects, The 2007 Revision Population Database. [Available online at http://esa.un.org/unup/.]

Voogt, J. A., and C. S. B. Grimmond, 2000: Modeling surface sensible heat flux using surface radiative temperatures in a simple urban area. *J. Climate Appl. Meteor.,* **39,** 1679–1699.

Willmott, C. J., 1981: On the validation of models. *Phys. Geogr.,* **2,** 184–194.

Wilson, M. F., A. Henderson-Sellers, R. E. Dickinson, and P. J. Kennedy, 1987: Sensitivity of the Biosphere–Atmosphere Transfer Scheme (BATS) to the inclusion of variable soil characteristics. *J. Climate Appl. Meteor.,* **26,** 341–362.

Zilitinkevich, S. S., 1995: Nonlocal turbulent transport: Pollution dispersion aspects of coherent structure of convective flows. *Air Pollution III—Volume I. Air Pollution Theory and Simulation,* H. Power, N. Moussiopoulos, and C. A. Brebbia, Eds., Computational Mechanics Publications, 53–60.

Appendix B: Initial results from Phase 2 of the international urban energy balance model comparison

INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **31**: 244–272 (2011)
Published online 7 October 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.2227

RMetS
Royal Meteorological Society

Initial results from Phase 2 of the international urban energy balance model comparison

C. S. B. Grimmond,[a]* M. Blackett,[a] M. J. Best,[b] J.-J. Baik,[c] S. E. Belcher,[d] J. Beringer,[e] S. I. Bohnenstengel,[d] I. Calmet,[f] F. Chen,[g] A. Coutts,[e] A. Dandou,[i] K. Fortuniak,[j] M. L. Gouvea,[a] R. Hamdi,[k] M. Hendry,[b] M. Kanda,[l] T. Kawai,[m] Y. Kawamoto,[n] H. Kondo,[o] E. S. Krayenhoff,[p] S.-H. Lee,[c] T. Loridan,[a] A. Martilli,[q] V. Masson,[r] S. Miao,[s] K. Oleson,[h] R. Ooka,[n] G. Pigeon,[r] A. Porson,[b,d] Y.-H. Ryu,[c] F. Salamanca,[q] G.J. Steeneveld,[t] M. Tombrou,[i] J. A. Voogt,[u] D. T. Young[a] and N. Zhang[v]

[a] *Department of Geography, King's College London, London WC2R 2LS, UK*
[b] *Met Office, FitzRoy Road, Exeter, EX1 3PB, UK*
[c] *School of Earth and Environmental Sciences, Seoul National University, Seoul 151-742, Republic of Korea*
[d] *Department of Meteorology, University of Reading, Reading, RG6 6BB, UK*
[e] *School of Geography and Environmental Science, Monash University, Melbourne, Vic, 3800, Australia*
[f] *Equipe Dynamique de l'Atmosphère Habitée Laboratoire de Mécanique des Fluides (UMR CNRS 6598) Ecole Centrale de Nantes, B.P. 92101, F-44321 NANTES Cedex 3, France*
[g] *Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado, 80307, USA*
[h] *Earth System Laboratory, National Center for Atmospheric Research, Boulder, Colorado, 80307, USA*
[i] *National and Kapodistrian University of Athens, Faculty of Physics, Department of Environmental Physics and Meteorology, Laboratory of Meteorology, Building Physics V, University Campus, 157 84 Athens, Greece*
[j] *Department of Meteorology and Climatology University of Lodz Narutowicza 88 Lodz Poland 90139*
[k] *Royal Meteorological Institute, Department II, section 3 Avenue Circulaire, 3, B-1180 Brussels, Belgium*
[l] *Department of International Development Engineering. Tokyo Institute of Technology,2-12-1-14-9, O-okayama, Megro-KU, Tokyo, Japan*
[m] *Research Center for Environmental Risk, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba-City, Ibaraki, 305-8506 Japan*
[n] *School of Engineering, The University of Tokyo, 7-3-1 Hongo, Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan*
[o] *Research Institute for Environmental Management Technology, National Institute of Advanced Industrial Sciecne and Technology,Tsukuba, Ibaraki, 305-8569, JAPAN*
[p] *Department of Geography, University of British Columbia, Vancouver, British Columbia, V6T 1Z2, Canada*
[q] *Department of Environment, CIEMAT, Madrid, 28040, Spain*
[r] *CNRM-GAME, Météo France/CNRS, Toulouse, 31057 Cedex 1, France*
[s] *Institute of Urban Meteorology, China Meteorological Administration, Beijing, 100089, China*
[t] *Meteorology and Air Quality Section, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands*
[u] *Department of Geography, University of Western Ontario, London ON N6A 5C2 Canada*
[v] *School of Atmospheric Sciences, Nanjing University 22 Hankou Road, Nanjing, 210093, China*

ABSTRACT: Urban land surface schemes have been developed to model the distinct features of the urban surface and the associated energy exchange processes. These models have been developed for a range of purposes and make different assumptions related to the inclusion and representation of the relevant processes. Here, the first results of Phase 2 from an international comparison project to evaluate 32 urban land surface schemes are presented. This is the first large-scale systematic evaluation of these models. In four stages, participants were given increasingly detailed information about an urban site for which urban fluxes were directly observed. At each stage, each group returned their models' calculated surface energy balance fluxes. Wide variations are evident in the performance of the models for individual fluxes. No individual model performs best for all fluxes. Providing additional information about the surface generally results in better performance. However, there is clear evidence that poor choice of parameter values can cause a large drop in performance for models that otherwise perform well. As many models do not perform well across all fluxes, there is need for caution in their application, and users should be aware of the implications for applications and decision making. Copyright © 2010 Royal Meteorological Society

KEY WORDS urban climate; energy balance; surface atmosphere exchanges; land surface modelling; sustainable cities; radiation; turbulent heat fluxes; evaporation

Received 29 March 2010; Revised 18 August 2010; Accepted 21 August 2010

1. Introduction

Land surface models (LSMs) parameterize energy exchanges between the surface and the atmosphere for a wide range of different land surface types (e.g. deciduous trees, coniferous trees, grasses, bare soil, and urban). They provide the lower boundary conditions (fluxes) to meso- and global-scale atmospheric models and are forced with meteorology from the overlying atmospheric model. A wide variety of approaches are taken to model the influence of the underlying land surface type. To

*Correspondence to: C. S. B. Grimmond, Department of Geography, King's College London, London WC2R 2LS, UK.
E-mail: sue.grimmond@kcl.ac.uk

model the exchanges for an urban environment, LSMs range from a relatively simple representation of the urban environment as an impervious slab to models that take into account the 3D geometry of buildings with varying heights and material characteristics (Grimmond *et al*., 2009, 2010). During the process of simplification inherent to modelling, urban LSM (ULSM) developers have also chosen whether or not, for example, to include turbulent latent heat and/or anthropogenic heat fluxes. Increasing complexity, however, comes at the cost of both greater computational requirements and the number of parameters requiring specification. As even the most complex models do not include the complete specifications of all exchange processes, of interest is what level of improvement in performance, if any, is obtained with increased complexity.

Previously ULSMs have been evaluated individually against observational datasets of fluxes (e.g. Grimmond and Oke, 2002; Masson *et al*., 2002; Dupont and Mestayer, 2006; Hamdi and Schayes, 2007; Krayenhoff and Voogt, 2007; Kawai *et al*., 2009; Loridan *et al*., 2010a, 2010b; Porson *et al*., 2010). Although providing useful insights, these studies lack a structure that facilitates robust intercomparison. Here the principles of the project for intercomparison of land surface parameterization schemes (PILPS) (Henderson-Sellers *et al*., 1993, 2003; Irranejad *et al*., 2003) are followed. This paper, the second in an international model comparison study (PILPS-urban), evaluates ULSM in a common and consistent manner. In the first paper (Grimmond *et al*., 2010), results from an evaluation that used a short dataset (14 days) for a known site were presented (hereafter called Phase 1). By knowing the site location, a modeller should be able to assign more appropriate parameter values. Here the results from a comparison of 32 urban LSM (Table I), which represent a range of approaches (Figure 1), are analysed for a longer dataset (16 months) with the participants initially not knowing the location of the site beyond its designation as urban (hereafter called Phase 2). The Phase 1 and 2 sites have very different land cover, most notably the amount of vegetation (less/more, respectively), and land use characteristics (industrial/residential, respectively). All participants in the second phase had to have completed Phase 1 (Grimmond *et al*., 2010); one model from Phase 1 is not part of Phase 2. Phase 2 was structured into four stages corresponding to the controlled release of information about the site to enable a comparison of the importance of the parameters for each of the models. Although each group is informed how their own model performs, each one is not told about individual performance of other models.

The objectives of this paper are

1. To evaluate the ability of ULSM, in general, to model urban energy balance fluxes when provided with varying degrees of information about the urban environment.
2. To evaluate the performance of models with similar characteristics and complexity.
3. To reveal opportunities for future improvement of ULSM.

The first objective aims to highlight what might be expected in terms of ULSM performance when modelling urban energy balance fluxes for an area when only limited information is available about the site. With a steady release of surface characteristics it is possible to assess what surface information is most critical for optimal model performance. With these results it is also possible to address the second objective, the results of which will aid users in assessing what type of modelling approach is most appropriate for further development or for a particular application.

2. Methodology

To participate in this comparison a model had to simulate urban energy balance fluxes from the forcing data provided (Table II). The urban energy balance for these purposes is defined as:

$$Q^* + Q_F = Q_H + Q_E + \Delta Q_S \quad (1)$$

where Q^* is the net all wave radiation flux density which consists of the incoming shortwave ($K_\downarrow$) and longwave ($L_\downarrow$) radiation, which was provided as part of the forcing data, and the outgoing shortwave ($K_\uparrow$) and longwave ($L_\uparrow$) radiation which have to be modelled as:

$$Q^* = (K_\downarrow - K_\uparrow) + (L_\downarrow - L_\uparrow) \quad (2)$$

The anthropogenic heat flux (Q_F) may be modelled, prescribed, or ignored. All models have to simulate the turbulent sensible heat flux (Q_H), but the turbulent latent heat flux (Q_E) is neglected by some (Figure 1). All models calculate the net storage heat flux (ΔQ_S). The advective flux is not included in the energy balance at this scale, although it does not mean that advection does not exist. The micro-scale advection should be included within the sub-grid surface flux parameterizations. At the meso-scale, the inter-grid variations would be resolved by the overlying atmospheric model. Here, the ULSM are run independently of any large-scale model (i.e. offline). This is to ensure that the model performance evaluates the ULSM and not any compensation occurring within a larger scale model. It also ensures that the atmospheric conditions are fixed and independent of larger scale model performance. Similarly, this comparison neither evaluates the facet or micro-scale energy balance fluxes nor the vertical profiles within the urban canopy of the mean meteorological variables that some of the models are capable of calculating. Here we discuss only the results for the directly observed fluxes, so the storage heat flux and anthropogenic heat flux are not discussed. These will be discussed in future papers.

To conduct this comparison, the principles of the PILPS are employed. At the completion of each of the

Table I. The number of versions of each model used in the comparison and number of groups using it.

Code	Model name	References	Versions	Groups
BEP02	Building effect parameterization	Martilli *et al.* (2002)	1	1
BEP_BEM08	BEP coupled with building energy model	Martilli *et al.* (2002); Salamanca *et al.* (2009, 2010); Salamanca and Martilli (2010)	1	1
CLMU	Community land model – urban	Oleson *et al.* (2008a, 2008b)	1	1
IISUCM	Institute of industrial science urban canopy model	Kawamoto and Ooka (2006, 2009a, 2009b)	1	1
JULES	Joint UK land environment simulator	Essery *et al.* (2003); Best (2005); Best *et al.* (2006)	4	2
LUMPS	Local-scale urban meteorological parameterization scheme	Grimmond and Oke (2002); Offerle *et al.* (2003); Loridan *et al.* (2010b)	2	1
NKUA	University of Athens model	Dandou *et al.* (2005)	1	1
MORUSES	Met Office reading urban surface exchange scheme	Harman *et al.* (2004a, 2004b); Porson *et al.* (2010)	2	1
MUCM	Multi-layer urban canopy model	Kondo and Liu (1998); Kondo *et al.* (2005)	1	1
NJU-UCM-S	Nanjing University urban canopy model-single layer	Masson (2000); Kusaka *et al.* (2001)	1	1
NJUC-UM-M	Nanjing University urban canopy model-multiple layer	Kondo *et al.* (2005), Kanda (2005a, 2005b)	1	1
NSLUCM/ NSLUCMK/ NSLUCM-WRF	Noah land surface model/single-layer urban canopy model	Kusaka *et al.* (2001); Chen *et al.* (2004); Loridan *et al.* (2010a)	3	3
SM2U	Soil Model for submesoscales (urbanized)	Dupont and Mestayer (2006); Dupont *et al.* (2006)	1	1
SNUUCM	Seoul National University urban canopy model	Ryu *et al.* (2009)	1	1
SRUM2/ SRUM4	Single column reading urban model tile version	Harman and Belcher (2006)	4	1
SUEB	Slab urban energy balance model	Fortuniak (2003); Fortuniak *et al.* (2004, 2005)	1	1
SUMM	Simple urban energy balance model for mesoscale simulation	Kanda *et al.* (2005a, 2005b); Kawai *et al.* (2007, 2009)	1	1
TEB	Town energy balance	Masson (2000); Masson *et al.* (2002); Lemonsu *et al.* (2004); Pigeon *et al.* (2008)	1	1
TEB-ml	Town energy balance with multi-layer option	Hamdi and Masson (2008); Masson and Seity (2009)	1	1
TUF2D	Temperatures of urban facets 2D	Krayenhoff and Voogt (2007)	1	1
TUF3D	Temperatures of urban facets 3D	Krayenhoff and Voogt (2007)	1	1
VUCM	Vegetated urban canopy model	Lee and Park (2008)	1	1

Note these are assigned anonymous numerical identifiers in the analysis.

four stages, additional site information was provided (Table II). In Stage 1, only the forcing data were provided along with knowledge that observations were measured at 6.25 times the mean roughness height (z_H) for an urban area. In later stages, more site information consisting of basic surface cover fractions (Stage 2), urban morphology (Stage 3), and characteristics of urban materials (Stage 4) was provided. From this information, further parameters could be derived by participants as necessary (Grimmond *et al.*, 2010). After the completion of each run, participants sent back the calculated fluxes and the parameter values used for their model runs.

The site selected for Phase 2 was chosen based on having (1) a year or more of data to allow seasonality to be incorporated into the modelling; (2) little previous use by modelling groups to test models; (3) an almost complete quality controlled dataset available (i.e. not just for certain meteorological conditions only); and (4) cooperation with those that were involved in the data collection to participate in PILPS-urban.

The Phase 2 observation site was in suburban (Preston) Melbourne, Australia (Coutts *et al.*, 2007a, 2007b). This location was concealed from participants until the completion of Stage 4 before which equivalent latitude

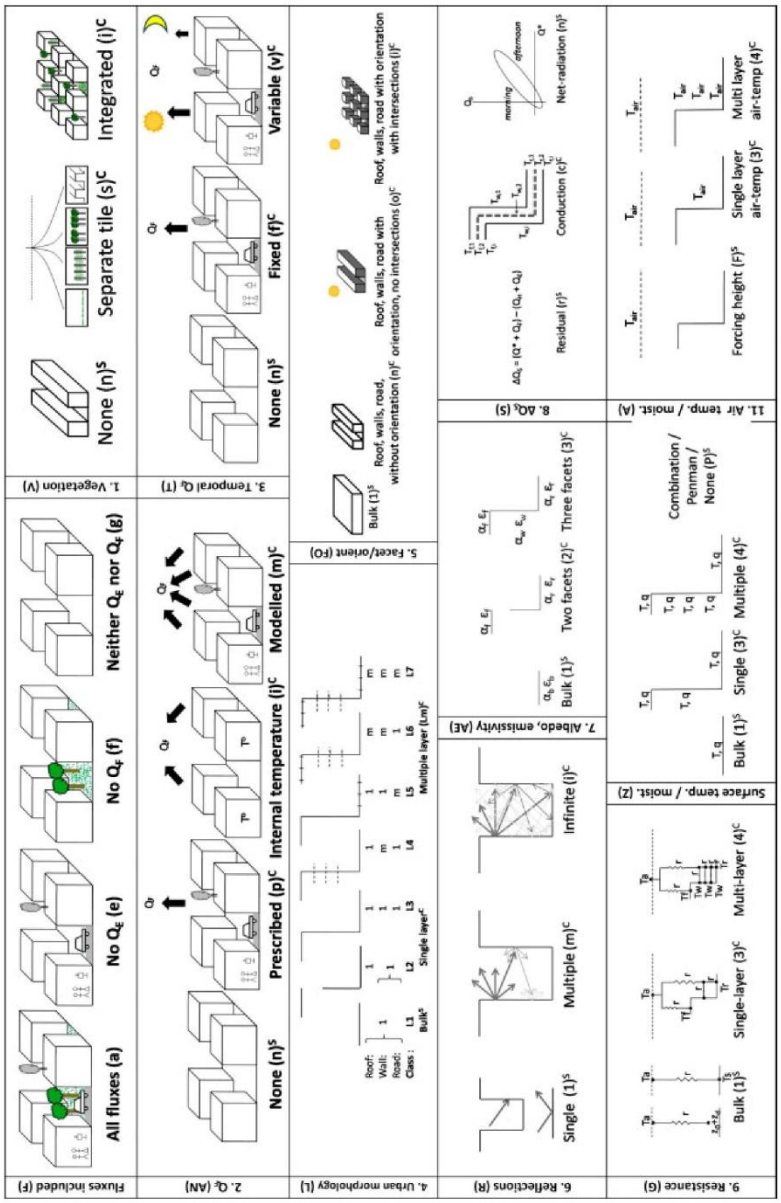

Figure 1. Model classifications with their individual characteristics. Each class characteristic is also classified as 'simpler' (S) or 'more complex' (C) (modified: Grimmond *et al.*, 2010). This figure is available in colour online at wileyonlinelibrary.com/journal/joc

Table II. Data provided at each stage.

	Category	Data provided		
Stage 1	Forcing data	K↓, L↓, air temperature, station pressure, specific humidity, wind components, rainfall		
	Site	Latitude*, Longitude*		
		Measurement height: 6.25 mean roughness height		
Stage 2	Plan area fraction	Pervious = 0.38		
		Impervious = 0.62		
Stage 3	Heights	Instrument height		40 m
		Roughness length for momentum		0.4 m
		Maximum height of roughness elements		12 m
		Mean building height		6.4 m
		Height to width ratio		0.42
		Mean wall to plan area ratio		0.4
	Plan area fraction	Surface cover	Fraction	Total
		Building	0.445	Impervious
		Concrete	0.045	0.62
		Road	0.130	
		Vegetation (excl. grass)	0.225	Pervious
		Grass	0.150	0.38
		Other (bare or pools)	0.005	
	Other	Urban climate zone = 5 Population density = 415.78 km^{-2}		

Stage 4: Material characteristics

	Wall	d	C_p	c	λ
	1	40.40	1008.5	1.25	0.61
	2	54.00	1456.3	1.40	0.43
	3	42.00	1010.0	0.0013[c]	0.024[d]
	4	12.50	837.0	0.67[a]	0.16[a]
	Roof				
	1	11.6	865.2	2.07	6.53
	2	50.00	965.3	0.0071	0.025
	3	40.00	1880.0	1.50[a]	0.23[a]
	4	12.50	837.0	0.67[a]	0.16[a]
	Road				
	1	28.75	912.79	1.14	1.17
	2	158.3[a]	840	1.05[a]	0.30[a]
	3	112.5[a]	840	1.29[a]	0.42[a]
	4	650.45[a]	801	1.43[b]	3.72[b]

ρ[a] brick = 1500; softwood = 560.5; hardwood = 800; concrete = 1822; asphalt = 2100; glass = 2535.7; asbestos cement building board = 1920; asbestos cement tiles = 1900; terracotta = 1700; metal = 7900; fibreglass = 60; air = 1.29; gypsum/plaster board = 800; coarse crushed rock = 1250; fine crushed rock = 1540; sandy loam = 1780

Site albedo = 0.15 | Site emissivity = 0.973

Stage 4: details of layers components for each facet

	Wall						Roof						Road					
1	Material	%	C_p	c^a	λ^a	d	Material	%	C_p	c^a	λ^a	d	Material	%	C_p	c^a	λ^a	d
	Brick	27.94	840.0	1.26	0.71	110	Metal	30	1105	8.73	72.00	3.0	Asphalt	75	920	1.10	1.2	35
	Softwood	59.43	1975.5	1.11	0.14	20	Tile: Concrete	40	837	1.52	1.10	16.5	Concrete	25	837.0	1.5	0.87	10
	Concrete	5.37	837.0	1.52	0.87	100	Tile: Terracotta	20	837	1.42	0.99	16.5						
	Asbestos cement	5.37	1005.0	1.93	0.58	8	Tile: Asbestos cement	10	945	1.79	0.55	8.0						
	Concrete/wood	1.13	1406.2	1.31	0.79	60												
	Metal	0.76	1105.0	8.73	72.00	3												
2	Hardwood	80	1880	1.50	0.23	40	Air	85	1010	0.0013	0.024	50	Coarse crushed rock					
	Brick	20	840.0	1.26	0.71	110	Fibreglass	15	712	0.04	0.03	50						
3	Insulation (air)						Wood						Fine crushed rock					
4	Gypsum/plaster board						Gypsum/plaster board						Soil (sandy loam)					

The exact latitude and longitude (*) were not known only an equivalent for solar zenith angle is used. The material characteristics provided at Stage 4 consisted of information for four layers for each facet (roof, wall, and road) that included: layer composition/material, layer width (d, mm), specific heat capacity (C_p, J kg^{-1} K^{-1}) and volumetric heat capacity (c, MJ m^{-3} K^{-1}) which are related through density (ρ, kg m^{-3}) and thermal conductivity (λ, W m^{-1} K^{-1}) as well as the site observed mean albedo and emissivity.

[a] Clarke *et al.* (1991).

[b] Ochsner *et al.* (2001).

[c] Engineering Toolbox (2005a).

[d] Engineering Toolbox (2005b).

and longitude for solar zenith angle were released. The radiative fluxes were measured using Kipp & Zonen CM 7B and CG4 radiometers. Temperature and relative humidity were measured using a Campbell Scientific Inc. (CSI) HMP45C sensor. Both were sampled at 1 Hz and averaged to 30 min. To evaluate the modelled fluxes, the outgoing radiation components and its net balance were determined from Equation (2). The turbulent sensible and latent heat fluxes were measured using the eddy covariance technique. A CSI CSAT3 3D sonic anemometer was used with a CSI krypton hygrometer (KH20, August 2003 to February 2004) or a LI-COR LI7500 open path infrared gas analyser (February 2004 to November 2004). They were sampled at 10 Hz and block averaged using a CSI CR23X datalogger. The fluxes were calculated for 30-min intervals (Coutts *et al.*, 2007a, 2007b). Diurnal and seasonal Q_F fluxes were estimated for the site, following Sailor and Lu (2004); the estimates include sources of Q_F from vehicles, buildings (from the consumption of electricity and natural gas), and human metabolism (Coutts *et al.*, 2007b). The storage heat flux was calculated as the residual to Equation (1). This approach has the inherent problem that it accumulates all the measurement errors and missing terms (e.g. horizontal advection ΔQ_A) in this flux (Grimmond and Oke, 1999; Offerle *et al.*, 2005). However, Offerle *et al.* (2005) and Roberts *et al.* (2006) obtained close correspondence between fluxes from detailed facet temperature measurements and local-scale residual estimates of ΔQ_S. It is important to recognize that for all observations measurement errors occur. The observed fluxes and the forcing data are not without errors which are systematic and unsystematic. Typical errors are related to the instruments and their calibration, the meteorological conditions under which the observations are taken (e.g. changing turbulence conditions, shading), the processing of the data, the representativeness of the turbulent and radiant footprint, and siting of the instruments (Offerle *et al.*, 2003; Lee *et al.*, 2004; Hollinger and Richardson, 2005; Dragoni *et al.*, 2007; Foken 2008). Hollinger and Richardson (2005) have demonstrated that the sizes of uncertainty increase as a function of net all wave radiation therefore increasing with the size of the flux; for the growing season for Q_H and Q_E, uncertainty increases roughly as 0.1 Q^* and 0.08 Q^* when $Q^* > 0$ W m^{-2} and above $\sim$10 W m^{-2} in an evergreen forest, respectively. They found no seasonality for Q_H errors but did for Q_E. Richardson *et al.* (2006) in an analysis of seven

sites, with a wider range of vegetation types, also found the error scales with the magnitude of the flux.

The forcing data consisted of 22 772 continuous 30-min intervals (474.5 days) from August 2003 to November 2004. Not all of the fluxes were available during all of these intervals, so here analysis is limited to the periods when all of the fluxes were measured. This gives 8865 intervals (38.9%) which were separated into two periods: the first 108 days and the last 365 days (8519 intervals when all fluxes were observed). The first period was to provide a spin-up, or initialization, period (the impact of this will be evaluated in a future paper). The post-initialization period allows for performance through an annual cycle to be evaluated.

Here, 32 different ULSMs are compared (Table I). The results are presented anonymously based on a randomly assigned unique model number. The models are grouped using a number of classifications based upon their characteristics (Figure 1) as described in Grimmond *et al.* (2010). To maintain anonymity, the number of models within each class had to be greater than three, thereby requiring some classes to be merged. Within each class, the approaches are categorized according to complexity (either simple or complex; Figure 1). Models are further categorized by their overall complexity depending on the number of 'complex' or 'simple' characteristics they possessed. The three groups are (1) 'complex' when all characteristics were complex (Cc), (2) 'medium' when the models possess one or two simple characteristics (Cm), and (c) 'simple' when they had three or more simple characteristics (Cs). Vegetation is not incorporated into this classification.

Comparison statistics reported here include root mean square error (RMSE), with both systematic ($RMSE_S$) and unsystematic ($RMSE_U$) components; the mean bias error (MBE); and the coefficient of determination (r^2). These are formally defined in Grimmond *et al.* (2010) from Willmott (1981) and Jacobson (1999). A larger systematic error typically indicates that the model has a problem in the model physics or parameter values, whereas a large unsystematic error is associated with the inability to cope with the variability in the observations which may be related to the 'randomness' of the conditions observed. Ideally, the systematic error would be the smaller of the two errors.

3. Results

3.1. Radiation comparison

To evaluate a model's ability to simulate radiative fluxes, the first aspect considered is whether there is closure in the radiation budget. Closure is assessed through comparison of the net all wave radiation (Q^*_{calc}) calculated from the two variables provided ($K_\downarrow$, $L_\downarrow$) and the two modelled variables ($K_\uparrow$, $L_\uparrow$) with the returned modelled Q^*_{mod}. No difference results in a coefficient of determination (r^2) of 1. At Stage 1, 15 of 32 models do not have a difference. In Stages 2/3/4, the number of models with $r^2 = 1$ is 13/16/13, respectively, but the total number of models that have $r^2 = 1$ at any stage is 18. Through four stages only ten models maintained no difference between Q^*_{calc} and Q^*_{mod}. If time periods with a difference of less than 1 W m^{-2} are considered (which includes one model with an r^2 of 0.999999), then Stages 1/2/3/4 have 16/14/16/13 models, respectively. These models are considered in the later analyses as being 'closed'. After this the r^2 values for Stage 1 range from 0.999991 to 0.0989 [*sic*]; with seven above 0.998, two more above 0.990, four more above 0.980, and two more greater than 0.870. The general groupings remain the same through the stages but the r^2 values do vary, except for the poorest models in Stages 1 and 2, which jump to greater than 0.998 at Stage 3.

Each modelling group which had a case of nonclosure was asked to determine the cause. The models without radiation balance closure problems are classified as P0 in the following analysis. Explanations for non-closure include (classified in analysis) not using the forcing data provided (P1), fluxes calculated independently (P1), timing issues (P3), day length (P3), spatial resolution (P3), and unknown (P4). In the first case, there are two different explanations: instead of using the individual 30-min interval forcing $K_\downarrow$ data, the daily peak observed $K_\downarrow$ was used and the other time periods for the day were obtained by assuming clear sky conditions, resulting in over-predicted $K_\downarrow$ and therefore Q^* (four cases, P1); and, the observed $L_\downarrow$ data were not used but modelled (one case, P1). In the second case, fluxes were calculated independently, the ULSMs calculate Q^* but for the purpose of this comparison, the radiative components have been calculated (three cases, P1) or there is an additional term in $L_\uparrow$ which is not incorporated into Q^* (one case, P4). In the third case, which relates to timing, the lack of closure is related to the 30-min forcing data being interpolated to a shorter time step for model calculations and then averaged back to the 30-min period for analysis (two cases, P3). This approach requires the forcing data to be interpolated which for $K_\downarrow$ may be questionable. For $L_\uparrow$, the approach depends on an emitted contribution from the surface temperature and a reflected part: $L_\uparrow(t) = (1 - \varepsilon)L_\downarrow(t - \delta t) + \varepsilon\sigma {T_S}^4(t)$. The surface temperature T_S depends on the energy received and has inertia. Alternatively, it is because $K_\uparrow$ is only calculated if the sun is above the horizon for the whole time interval (one case, P3), thereby impacting the day length. The fifth case of spatial resolution (two cases, P3) is related to an underestimation of the total sky view factor (all model patches sum to less than 1.0) that arises in the process of rasterizing the surface within the model. The affected models then absorb slightly too much or too little diffuse solar or longwave radiation. The final case is where there are problems which the modelling groups have not been able to determine, leading to the imbalance (three cases, P4).

3.1.1. *Outgoing shortwave radiation*

The performance of each model, with respect to outgoing shortwave radiation ($K_\uparrow$), is shown in Figure 2 based on

RMSE; models that do not have closure are indicated. For this upwelling solar flux, only daytime fluxes are analysed. This gives 4266 × 30-min periods for comparison. The mean observed flux is 54.2 W m^{-2}. The Stage 1 $K_{\uparrow}$ mean RMSE for all ($N = 32$)/($N = 31$ models – as model 17 did not complete all stages)/not-closed/closed are 28/17/42/15 W m^{-2}, respectively, but the large difference is because of one model (17) which does not have closure. The mean RMSE for all 32 models by stage is generally larger than the median (Figure 2) because the mean is impacted by two poorly performing models, one of which did not complete Stage 4.

Considering all 32 models, as increasing information was provided (Stages 1–4) there was an improvement at each stage in mean but not in the median RMSE. The median RMSE improves from Stage 1 to 2 and again between Stage 3 and 4 (Figure 2). Of the 16/32 models with an improved RMSE from Stage 1 to Stage 2, 7/16 improved from Stage 2 to 3; and 2/7 of those improved from Stage 3 to 4. Thus, only two models had a reduction in RMSE at each stage. At Stage 2, improvement is associated with the fraction of vegetation to built areas becoming known (Table II). This fraction provides for the more realistic assignment of 'urban' and 'vegetated' albedos within the models. However, RMSE for five models became poorer. In Stages 2 and 3, a total of 14 models reduced (and 14 models increased) their RMSE and 13 in Stages 3 and 4 (and 4 increased). At Stage 3, more detailed information was provided about the surface fractions and heights. For the urban fraction it was now possible to distinguish the road and roof fractions correctly, in addition to knowing the wall heights. In the pervious fraction, grass could be distinguished from other vegetation. As expected, the largest overall improvement in $K_{\uparrow}$ based on the mean and median RMSE occurred at Stage 4 when the site observed albedo was provided (Figure 2).

The relative ordering of models in terms of performance remains relatively similar for all stages for $K_{\uparrow}$ with the same three models performing in the top three for all stages (Figure 2). Similarly, the poorest performing models, with slight reordering, remain the same for the four stages. But there are some notable changes for individual models between stages; e.g. model 22 does very well in Stages 1 and 2, then in Stage 3 the performance is much poorer but then returns to very good performance for $K_{\uparrow}$ in Stage 4. This demonstrates the importance not only of the model physics but also of the user's choice of parameter values, which can significantly influence the outcome. For Stages 1–3, there is a larger median systematic error ($RMSE_S$) than unsystematic error ($RMSE_U$), even when excluding model 17, but not for Stage 4 (Figure 2), suggesting that the additional surface information is important for improving the model performance. In Stage 4, once information about the albedo is available, 80% of the models have an $RMSE_U$ that is greater than the $RMSE_S$. The shading of the bars distinguishes the models complexity (C) among simple (s, yellow, light grey), medium (m, blue, medium grey), and complex (c, crimson, dark grey) (see Section 2 for definition). It can be seen that the three model types are distributed across the range of model performances, with all three occurring in the first and last five at Stage 1. By Stage 4, the Cc models are all in the middle group, (except a Cc model has dropped out). At Stage 4, the majority of the Cs models are doing well but the poorest performing model belongs to that group.

The effective albedo (α_{eff}) used in the models can be determined from $K_{\uparrow mod}/K_{\downarrow obs}$. Here this value is investigated at two times of the year (June 21 and December 21) at 13 : 00 h. These two times will have maximum and minimum amount of midday shadow. The range of values at Stage 1 is from 0.08 to 0.28 (except for two extreme outliers). The best performing model had an α_{eff} of 0.15, which was the same as the observed value provided at Stage 4, on both dates. The December 21 range of values were 4 (3) cases <0.1 (or >0.2); 3 (4) cases that were 0.10–0.125 (0.175–0.20); and 16 cases with an α_{eff} within 0.125–0.175, of which 11 have the lowest RMSE for $K_{\uparrow}$. For June 21, there was a similar distribution. The slightly higher α_{eff} (0.175–0.18) values are associated with the next best cohort in terms of RMSE performance.

The average cohort MBE is strongly influenced by the poorest performing models (Figure 2). The models have both positive and negative biases across the range which results in a net small negative bias (−4 W m^{-2} excluding model 17) for Stage 1. The median MBE has a large improvement from Stage 1 to 2 but after that remains almost constant at −1 W m^{-2}. At Stage 4, the Cm models which perform least well all have a negative bias, whereas the poor Cs models have both positive and negative MBE.

On a normalized Taylor (2001) plot, where the ideal model performance is indicated by the open circle at 1.0, 1.0, 0.0 (Figure 3), the correlation coefficient (polar) and normalized standard deviation (y-axis) and normalized RMSE (inner cirles) are shown. Except for one model the correlation coefficient is better than 0.8; for the majority of the models it is better than 0.9; and for many better than 0.99. One can track the impact of the additional information for the individual models; e.g. model 44 (medium complexity so blue with a symbol of a plus sign within a circle shown in Figure 3) in Stage 1 had a correlation of ∼0.85 which improved in Stage 2 to ∼0.91 and improved again in Stage 4 to ∼0.95. Between Stages 2 and 3 there is a very minor change in correlation. In addition, one can see that there is an improvement in the normalized RMSE from greater than 0.5 to 0.4 to less than 0.4 (ideal is 0.0); and improvement of the normalized standard deviation from 0.62–0.73 to 0.74–0.88 (ideal is 1.0). For model 46 (same symbol but simple complexity) one can see that the model does not systematically improve.

Ensemble modelling, where the mean result from a number of different models is reported, is now used quite

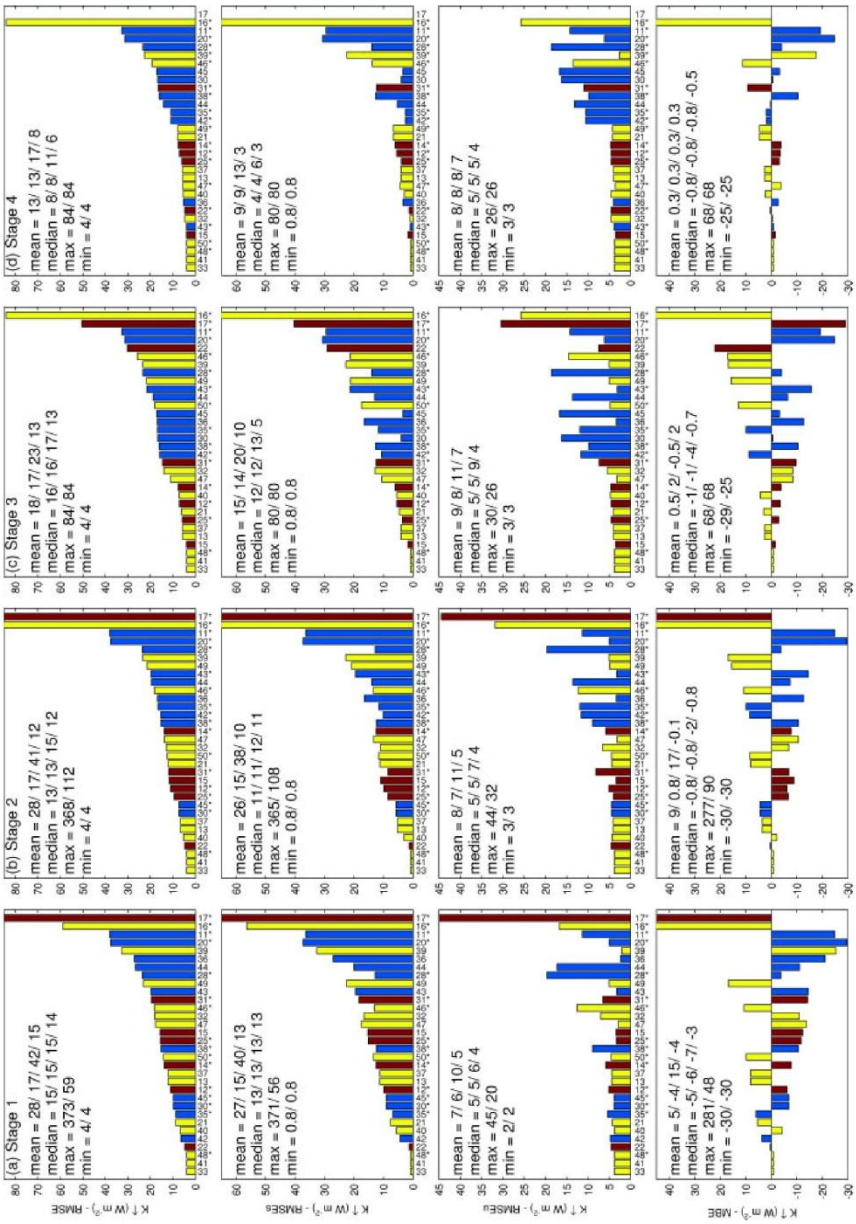

Figure 2. Model performance for each of the four stages (columns), for the last 12 months for outgoing shortwave radiation ($K_{\uparrow}$) (daytime only). The models are ranked based on RMSE for each stage with the systematic and unsystematic RMSE and MBE shown in the same order for each model. The overall statistics (mean, median, maximum, minimum) are given for the 32 models with and without model 17 ($N = 31$) for each stage in each figure. For the mean and median the statistics are also given for those models which do not have closure and do have closure of the radiation balance. The models which do not have radiative closure (see text) are indicated with a '*' (a–d). The shading of the bars distinguishes the models overall complexity (C) between simple (s, yellow, light grey), medium (m, blue, medium grey), and complex (c, crimson, dark grey) (see Section 2 for definition). The mean observed flux for this period was 54.2 W m^{-2}. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

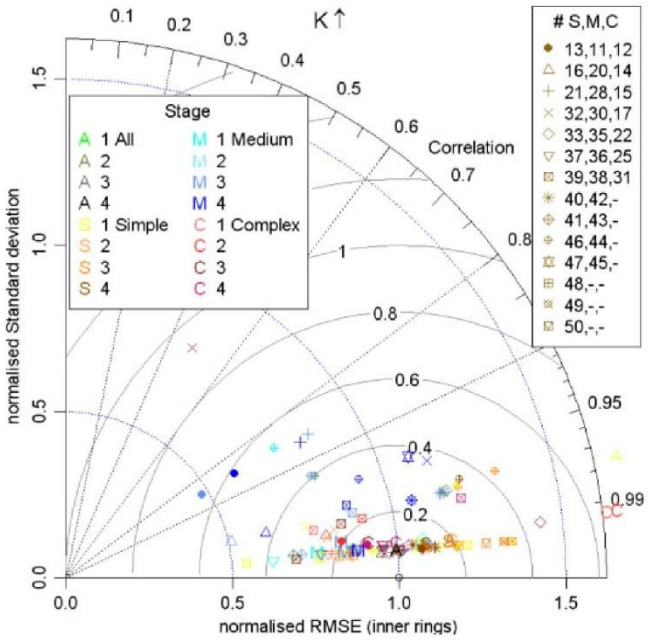

Figure 3. Normalized Taylor plot for the four stages, for the last 12 months for outgoing shortwave radiation ($K_{\uparrow}$) (daytime only). Taylor plots have the correlation coefficient on the polar axis, the normalized standard deviation on the radial *y*-axis and the normalized RMSE (*x*-axis) on the internal circular axes (Taylor, 2001). Performance for each model (symbol, colour indicates complexity and stage) and the ensemble results by complexity (letter) and stage (colour) are shown. Legend symbols are shown for simple Stage 4 colour. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

extensively in the climate community (e.g. Gillett *et al.*, 2002). Here we consider the model performance of four different ensembles, three based on complexity [simple with 14 models, medium with 11, and complex with 7 (or 6 when model 17 drops out in Stage 4)] and the fourth is when all of the models are included [32 (Stages 1–3) or 31 (Stage 4)]. In Figure 3, these are shown for each stage. For the simple models, the correlation remains approximately constant, but there is an improvement in both the normalized standard deviation and normalized RMSE in the ensemble performance with stage. This is also the case for the medium and complex models. However, the ensemble performance of the complex models is clearly strongly influenced by the outlier model (17), which is beyond the plot boundaries, in Stages 1 and 2. At Stage 4 the ensemble performance is best when all (A) models are used but this is only slightly better than the ensemble mean performance of the complex models; the simple models' ensemble mean is slightly better than the medium complexity models.

The characteristics used to classify the models (Figure 1) include some that are directly related to radiative modelling. When the model results are grouped by these characteristics (Figure 4), we can determine if particular approaches result in better performance. In several classes, there is a clear separation in the mean performance associated with modelling $K_{\uparrow}$. However, in many cases the change in the mean is caused by one model's performance so the median is more robust as a measure of central location within the data. To maintain anonymity, each set of results plotted was required to have four or more results. This means that some classes are amalgamated. For each characteristic at each stage a box-plot of the RMSE gives the interquartile range (IQR), the individual models are plotted as dots, the median as a square, and the mean as a circle. Below each box the stage, the classification type, the characteristic with the class, then the number of models, the median, and the mean appear. For example, Figure 4(a) 1-Vn/11/14/17 indicates that for Stage 1 when the models are classified based on their approach to vegetation (V), there were 11 models that did not include it (n) with a median RMSE of 14 W m^{-2} and a mean of 17 W m^{-2}.

The first characteristic considered is whether the model integrates vegetation with the urban tile (Vi) rather than treating it separately (Vs) or not including it at all (Vn). For the Vi models there is a clear improvement in all four stages (Figure 4(a)). By Stage 3, the Vi models have a median RMSE of <4 W m^{-2} which is the smallest value. From Stage 2, when more models included vegetation (Vs models increase in number at

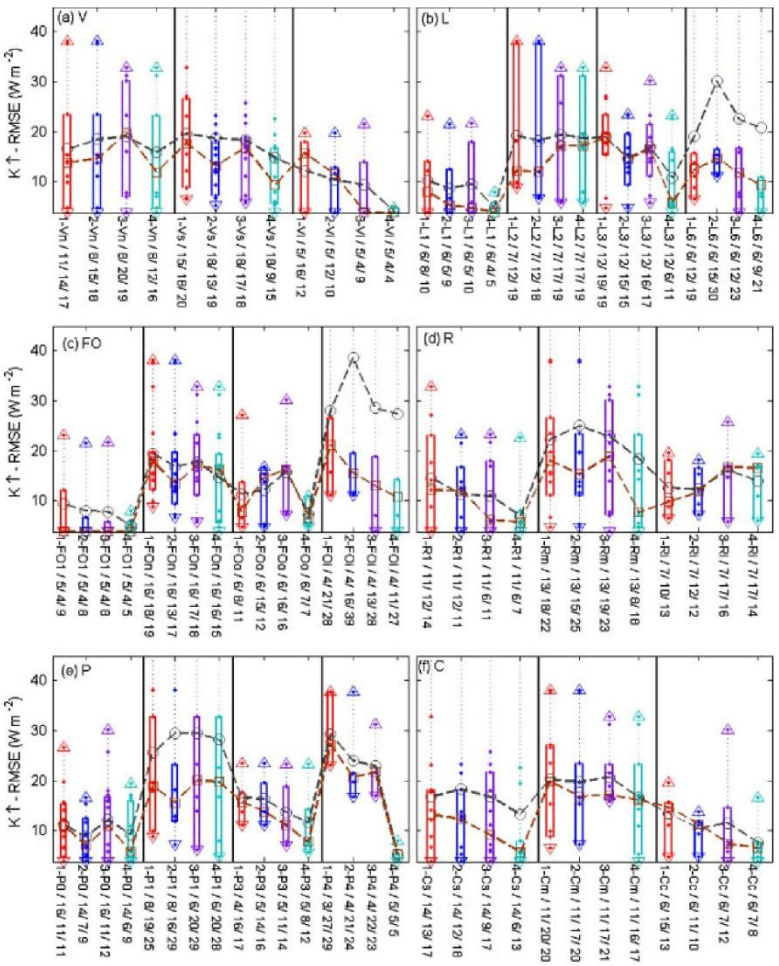

Figure 4. Model RMSE performance for the four stages, using the last 12 months for outgoing shortwave radiation ($K_{\uparrow}$) (daytime only) for the classes by approach taken (see Figure 1 for code interpretation or text): (a) vegetation (V), (b) urban morphology (L), (c) facets and orientation (FO), (d) reflections (R), (e) radiative closure performance (P), (f) complexity (C). Individual models are shown by the points, maximum and minimum by the triangles and the IQR by the box. Note the plots are cut-off at 0.40 of the maximum and the statistics are for $N = 31$ models (excludes 17). The circles are the mean of the cohort and the square is the median. The number of models, median, and mean are given for each. See text for further details. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

the expense of Vn) the model cohorts retained the same ordering Vn, Vs, and Vi (decreasing median RMSE), but both Vn and Vs median performances deteriorate slightly in Stage 3. We can conclude that accounting for vegetation is important which is consistent with the conclusions from Phase 1 (Grimmond *et al.*, 2010).

Urban morphology (L) is specified using seven different approaches; from a slab surface (L1) to single-layer models (L2 – two components, L3 – three facets) and multi-layer (L4–7) models. The multi-layer models (L4–7) have different aspects of the surface that are treated in more detail (Figure 1) which leads to small numbers in each class. In this paper, these have been grouped together and labelled L6. This group has by far the largest mean RMSE because of one outlier (Figure 4(b)). The median performance for the simplest slab models (L1) improves at each stage and has the lowest median RMSE at Stage 4. For the other classes, there is not a consistent trend between stages; and for the L2 models the Stage 3 and 4 results have a higher median although reduced range, maximum and minimum than the earlier stages. The L3 models have second best median RMSE at Stage 4. Note for this characteristic that there

is no change in the model numbers per cohort between stages.

The approach to surface geometry with respect to whether the surface explicitly includes shaded surfaces or not (FO) has distinct differences between groups (Figure 4(c)). The simplest case, where the surface has a bulk geometry (FO1), has the lowest RMSE at all stages. It has a median RMSE of 4 W m^{-2} for all stages; however, the IQR decreases indicating more similar results. The most complex approach, which has both shading and intersections (FOi), has a systematic decrease in median RMSE at each stage, but at Stage 4 it is 11 W m^{-2}. This is greater than for models that take shading into account but have no intersection (i.e. have infinitely long canyons) (FOo) which have a median RMSE at Stage 4 of 7 W m^{-2}. The FOi models are clearly benefitting from the additional information provided, such as the wall height and built fraction provided at Stage 3. Both the FOo models and those that have an infinitely long canyon but do not account for shaded areas (FOn) have varying behaviour between stages; neither shows a continuous or significant improvement. The latter have the larger median RMSE at Stage 4 (16 W m^{-2}). The changing geometry influences the complexity of the modelling significantly with the simplest FO1 requiring considerably less computer resources than the more complete FOi which is theoretically much more realistic if within canyon information is required. Note, however, that the ability to model in-canyon information is not actually evaluated here.

Not only may the surface morphology description be different, but the approach taken to model reflections (R) also varies from those that include single (R1), multiple (Rm), or infinite reflections (Ri). The simplest (R1), unlike the other two approaches, has a systematic improvement in the median RMSE with stage (Figure 4(d)). By Stage 4, the median RMSE of 6 W m^{-2} is the smallest of the three approaches. The Rm approach, although it has a large scatter, shows a net improvement by Stage 4 (median RMSE = 8 W m^{-2}). The Ri group (median RMSE = 17 W m^{-2}) actually deteriorates through stages. So the simplest group consistently is the best performing and benefits from the additional information provided.

The albedo and emissivity (AE) classification distinguishes the amount of parameter information that is required by the models. The simplest case requires one bulk value (AE1) and so has a similar behaviour to FO1 and L1 (not shown). Significant improvement for these models at Stage 4 is a simple consequence of model formulation. Prior to Stage 4 albedo was assumed, but in Stage 4 for some models $K_{\uparrow}$ is just the product of two given values: site albedo and $K_{\downarrow}$. Models also can require two values (per parameter) typically associated with two facets (AE2) or three or more values (AE3). The median RMSE is lowest for the AE1 group and largest for AE2 (median RMSE at Stage 4 is 4 and 20 W m^{-2}, respectively). The vast majority of the models (22) require at least three values (AE3) for which the median RMSE by Stage 4 is 9 W m^{-2}; a net improvement from Stage 1. However this group, like the Rm, continue to have a wide range of values for the individual models.

The models that do not have a problem with net radiation balance closure (P0) have the smallest median RMSE at each Stage (Figure 4(e)). Their IQR does not have the smallest spread but the minimum values are lowest, and except for Stage 4, the 75 percentile is the lowest. The P3 (time and space resolution issues) and P4 models (unknown) have a systematic improvement with stage. At Stage 4, the median RMSE is 6/20/8/5 W m^{-2} for the P0/P1/P3/P4 models, respectively. The P1 models that have problems calculating a component of the radiative balance or did not use the forcing data for individual time intervals perform poorly throughout.

For all three model complexities, there are steady improvements in performance as additional information is provided (Figure 4(f)). The simplest and most complex (Cs, Cc) have a larger overall improvement than the Cm models with additional surface information. The Cs models have a slightly better median (6 rather than 7 W m^{-2}) but the mean is better for the Cc models (8 W m^{-2}).

Overall $K_{\uparrow}$ is modelled well and the provision of additional information about the surface does result in better performance. The models that perform best, for individual characteristics, are those that are the simplest as they can be assigned one parameter that is close to the observed value. The inclusion of vegetation is important to the performance. Based on overall complexity the simplest and the most complex models have similar results. The models that have net radiation closure perform better generally. The poorest performing cohort overall (P1) at Stage 4 does not have radiative closure and either did not make use of the individual time interval data and/or calculated the fluxes independently.

3.1.2. Outgoing longwave radiation

A combination of parameter information and flux calculations impact surface temperatures and hence the outgoing longwave radiation flux ($L_{\uparrow}$). Thus, the modelling of day- and night-time $L_{\uparrow}$ is more complex than modelling $K_{\uparrow}$ because of the relation between surface temperature, sensible heat, and storage heat fluxes, as well as $L_{\uparrow}$ itself. This means that, unlike the $K_{\uparrow}$ case, when additional information is provided more related parameters may be influenced.

For $L_{\uparrow}$, the median RMSE for the 32 models from Stages 1/2/3/4 are 16/13/14/17 W m^{-2}, respectively (Figure 5). Overall, 18 models improved from Stage 1 to 2, 11 from Stage 2 to 3, and 8 from Stage 3 to 4. Of the 32 models, only two improved across all the stages but eight improved in three consecutive stages. The largest improvement for an individual model was from Stage 2 to 3 with a greater than 20 W m^{-2} decrease in RMSE. The model performance from Stages 3 to 4, despite now having the most information about the site (Table II), suffered the largest loss of performance with 23 models

having an increase in RMSE. This relates to the trade-off that is made in parameter values. The largest individual performance deterioration also occurred between Stages 3 and 4 (increase of >35 W m^{-2} in the RMSE). There was one model that deteriorated across all four stages.

The models that close the radiation balance generally have better performance (e.g. smaller median RMSE) but that is not the case in Stage 1. At all stages, the models have a larger mean $RMSE_S$ than $RMSE_U$ but by Stage 3 and 4 the median $RMSE_U$ is slightly larger (Figure 5), suggesting that the model parameter information is appropriate for most of the models. In terms of the MBE more models have a positive bias rather than negative, but the two (one at Stage 4) models which perform least well have a large negative bias. The median MBE remains at about 8 W m^{-2} across all four stages.

The overall range of RMSE is smaller for $L_\uparrow$ than $K_\uparrow$ but the best performing model for $L_\uparrow$ has a larger RMSE than the best model for $K_\uparrow$. From comparison of the normalized Taylor plots (Figures 3 and 6), it is clear that the correlation is generally poorer for $L_\uparrow$. The mean $L_\uparrow$ flux is larger, but the diurnal range is smaller, than $K_\uparrow$. As with $K_\uparrow$, one (although different to $K_\uparrow$) model performs best across almost all stages (based on RMSE) and shows very little improvement with additional information being provided. This again is a simple model (Cs). The poorest performing model (excluding Model 17) does improve slightly with additional site information but still has a larger $RMSE_S$ than $RMSE_U$, suggesting that the model could be improved further. This differs from the next least well-performing model which has a larger $RMSE_U$ and a small positive MBE.

The three classes of complexity are scattered across the range of performance. However, again the best and poorest models are simple (Cs). In general, the simpler models are grouped in the middle or poorer end by Stage 4, whereas many of the Cm models are amongst the best. Unlike for $K_\uparrow$ the ensemble mean performance of the models does not improve with stage (Figure 6). At Stage 4 for all four ensembles all three measures have deteriorated. There is one model that is clearly performing better than the ensemble (but this is not the model with the lowest RMSE) pre-Stage 4. From the Taylor plot the best performing ensemble is the medium complexity but the four ensembles are clustered (and have moved together as a cluster between stages).

There is no model class that is better than the others. In most cases the model cohorts show poorest performance for all classes in Stage 4. For example (Figure 7), at Stage 4 the IQR is greater than in Stage 3 for all the approaches taken for vegetation (V); treatment of the urban morphology (L) has a drop in performance for each cohort in Stage 4, with the more complex models (L6) having the largest increase in median RMSE. There is very little change between stages in the other L classes. A similar result is obtained for the facet and orientation characteristics (FO) with no cohort improving across all four stages. One class (FOo) has a 6 W m^{-2} increase in median RMSE. For R and AE, similar results are obtained.

The models that have radiative closure (P0) have a median RMSE of 15 W m^{-2} at Stage 1 and 4. At Stage 4, the P0 cohort has the lowest median but this is not the case for all stages. For those without closure, the Stage 4 median is larger in all cases than Stage 1. For all P classes, Stage 2 was when the median RMSE was smallest.

The modelling of $L_\uparrow$ initially has the same size median RMSE as $K_\uparrow$ but not the general improvement with additional information (or progressive stage). This is seen consistently across all the classes of model types. In most cases, the Stage 4 results are poorer and have a larger IQR. At Stage 4, the best performing modelling approaches (lowest median RMSE) have the Vi, L3, FO1, Ri, AE1 and Cc characteristics. As was demonstrated previously (Grimmond *et al*., 2010, Fig. 3), no single model has all these characteristics.

The models perform generally better at night than over the 24 h period (mean observed flux day = 410.14 W m^{-2} and night = 368.98 W m^{-2}). At night, the median RMSE for Stages 1/2/3/4 are 12/11/10/12 W m^{-2} and the median MBE are 8/7/2/−0.2 W m^{-2}. At Stage 4, the best performing (median RMSE W m^{-2}) models have Vn (13)/L2 (10)/FOn (11)/Rm (11)/AE2 (10) characteristics. Notably there is no difference between Cs/Cm/Cc models; they all have a median RMSE of 12 W m^{-2}. The daytime, as expected, is poorer with median RMSE for Stages 1/2/3/4 of 18/14/16/20 W m^{-2} and the median MBE are 9/7/9/12 W m^{-2}. At Stage 4, the best performing (median RMSE W m^{-2}) models have Vi(16)/L2 (17)/FOi (18)/Ri (15)/AE1 and AE3 (20)/Cc (15) characteristics. Thus, the characteristics that result in the lowest median RMSE change with time of day so there is not a clear choice, although the differences in the errors are small.

The models that do not have radiative closure occur across the complete spectrum of model performance for all time periods. The daytime median RMSE for P0 models improves from Stage 1 to 4 from 18 to 16 W m^{-2} but the Stage 2 result is the best for P0/P3/P4 models. For P1 models, the best performance is Stage 3 (15 W m^{-2}) but at Stage 4 the median RMSE is the poorest (26 W m^{-2}). At night the median RMSE for P0 models is 11 W m^{-2} at all stages (but deteriorating). The best performance is Stage 3/2/4 for P1/3/4 models.

Overall $L_\uparrow$ is not as well modelled as $K_\uparrow$. The daytime, when the mean flux is larger, has the larger median RMSE. The models generally improve when information about the pervious/impervious fraction is provided but generally did not improve when further details about heights and surface fractions were provided. Most models deteriorated when they were provided with details of the building materials typically back to Stage 1 performance but in many cases even poorer. Given both the wide range of materials that are in urban areas and the

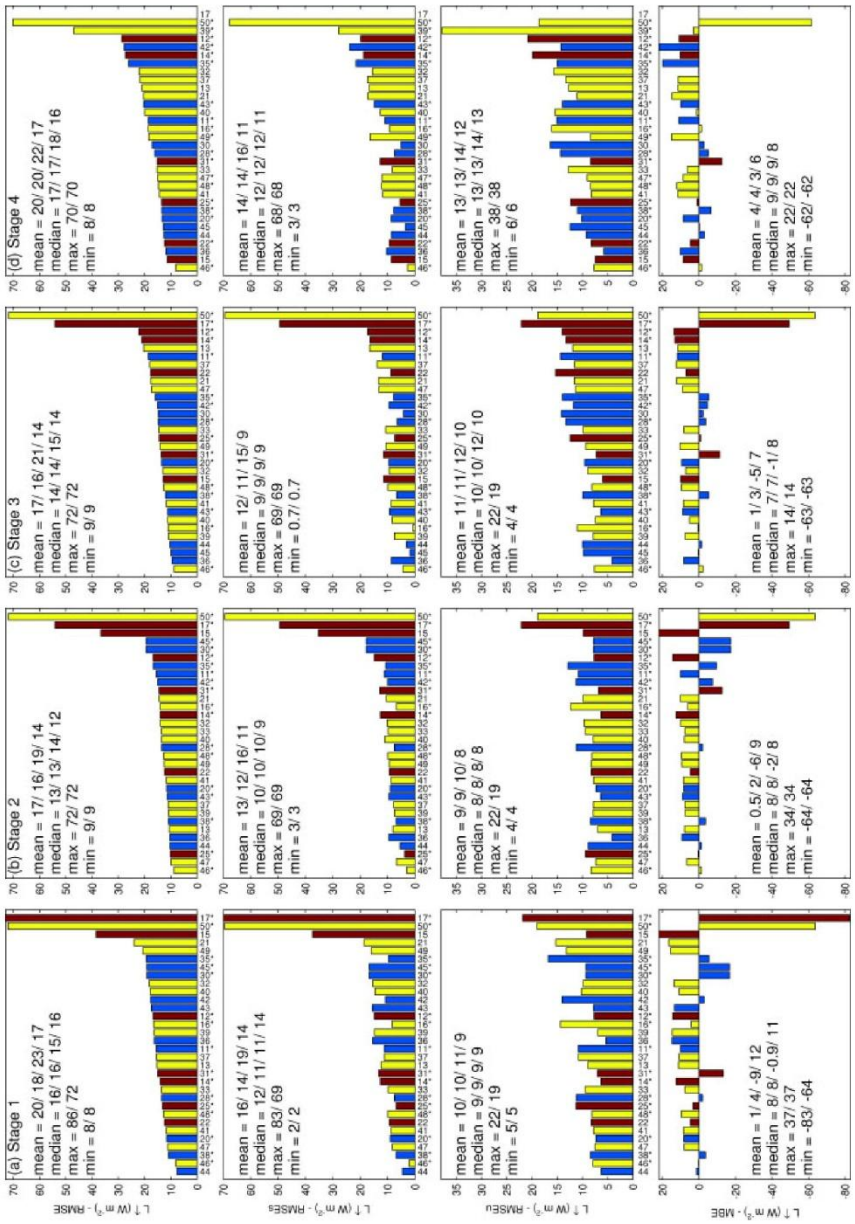

Figure 5. As for Figure 2 but for outgoing longwave radiation (L↑) for all hours. The mean observed flux for this period was 389.6 W m^{-2}. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

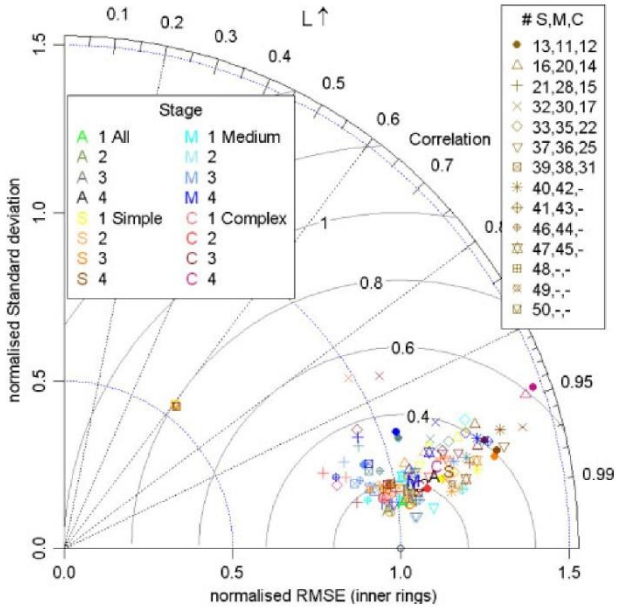

Figure 6. As for Figure 3 but for outgoing longwave radiation (L↑) for all hours. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

associated wide range of values for individual material types, the difficulty of deciding what the appropriate values should be suggests that until there is a way to obtain realistic values for actual sites specification of materials may not be worth the effort required to obtain the information. Here we contacted a large number of people associated with the building and planning design plus materials suppliers (see 'Acknowledgements') to allow us to provide the data in Table II.

3.1.3. *Net all wave radiation*

Figure 8 shows the ranked performance of the models based on RMSE of net all wave radiation (Q^*), with the lack of radiative closure indicated. It can be seen from Figures 2, 5, and 8 that models which do not have closure are distributed from the best performing to the poorest performing for all three radiative fluxes evaluated, but are mainly the poorest performing for Q^*. For Stage 1 the mean RMSE for all models is 29 W m^{-2} for Q^* or 28 W m^{-2} when the model with poorest closure (r^2 of 0.0989) is removed because it did not complete all four stages. However, this model is not the poorest performing for Q^* but is for $K_{\uparrow}$ and $L_{\uparrow}$ at Stage 1 (Figures 2 and 5). Models that have radiative closure generally perform better over all stages for Q^* than those that do not; on average having a mean RMSE 20 W m^{-2} smaller. However, closure of the radiation balance is not a good measure of ability to calculate a particular flux. Comparing the performance of the components to the net all wave radiation shows a clear re-ranking between fluxes. Notably those that perform poorly for an individual component flux are not the poorest for Q^* (Figures 2, 5, and 8). This means that the application that the model is being used for is important; for example, when assessing a mitigation strategy's impact (such as changing the albedo of the materials on the change in radiative fluxes and temperatures) an ULSM may be modelling the most directly impacted flux well, but not able to model the other fluxes (or vice versa).

There were 14 models which showed a reduction in RMSE from Stage 1 to 2; of these five had a further improvement at Stage 3; and two of these improved again at Stage 4. However, in the opposite situation there are eight models whose RMSE increased from Stage 1 to 2; of which five had a further increase at Stage 3 and four had a further drop in performance at Stage 4.

The overall performance for Q^* does not vary much between stages though, with the mean RMSE being approximately 30 W m^{-2} at Stage 4, which is slightly larger than in the earlier stages. Also at Stage 4 models that do have closure of the radiation balance have a smaller mean and median RMSE (both 18 W m^{-2}, Figure 8). At Stage 4, however, these models have a slightly larger $RMSE_S$ than $RMSE_U$ suggesting that an improvement could still be made in the physics or parameter specification but this is not the case for both K↑ and L↑. The models generally have a negative MBE (Figure 8, Stage 4 median −6 W m^{-2}). The models with

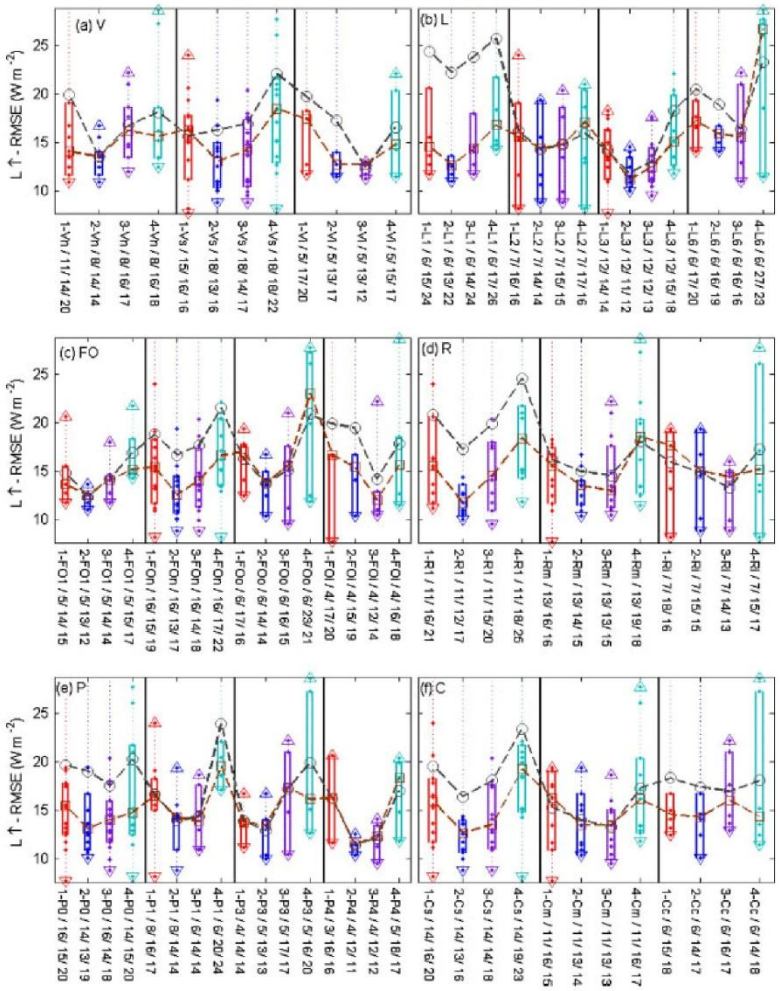

Figure 7. As for Figure 4 but for outgoing longwave radiation (L↑) for all hours. Note plots are cut-off at 0.40 of the maximum. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

the largest absolute MBE are both positive and negative (Figure 8).

The best and poorest performing models at all stages are of medium complexity (Cm). At Stage 1 at both ends of the performance spectrum we have models from the three levels of complexity. By Stage 4 the more complex models have generally improved with three of the six (remember model 17 no longer appears) best performing models. Cm models are grouped more at the end with poor performance.

From the Taylor plot (Figure 9) it is clear, except for three models, all do an excellent job of modelling Q^*. There is a very tight cluster around (but not on) the ideal point. This performance is clearly better than for the separate radiative fluxes. Although this is good, this does suggest that there is some compensation occurring within the individual fluxes which may not be physically correct. As noted previously this result suggests that caution is needed when using the models to account for changing radiative characteristics. For the ensemble performance the medium complexity models are poorer than the other three. The best are the simple and complex models with slightly poorer performance from the 'all' ensemble.

The models that do not account for vegetation (Vn) show a steady decline in performance across all stages (Figure 10(a)). In contrast, there is no strong evidence

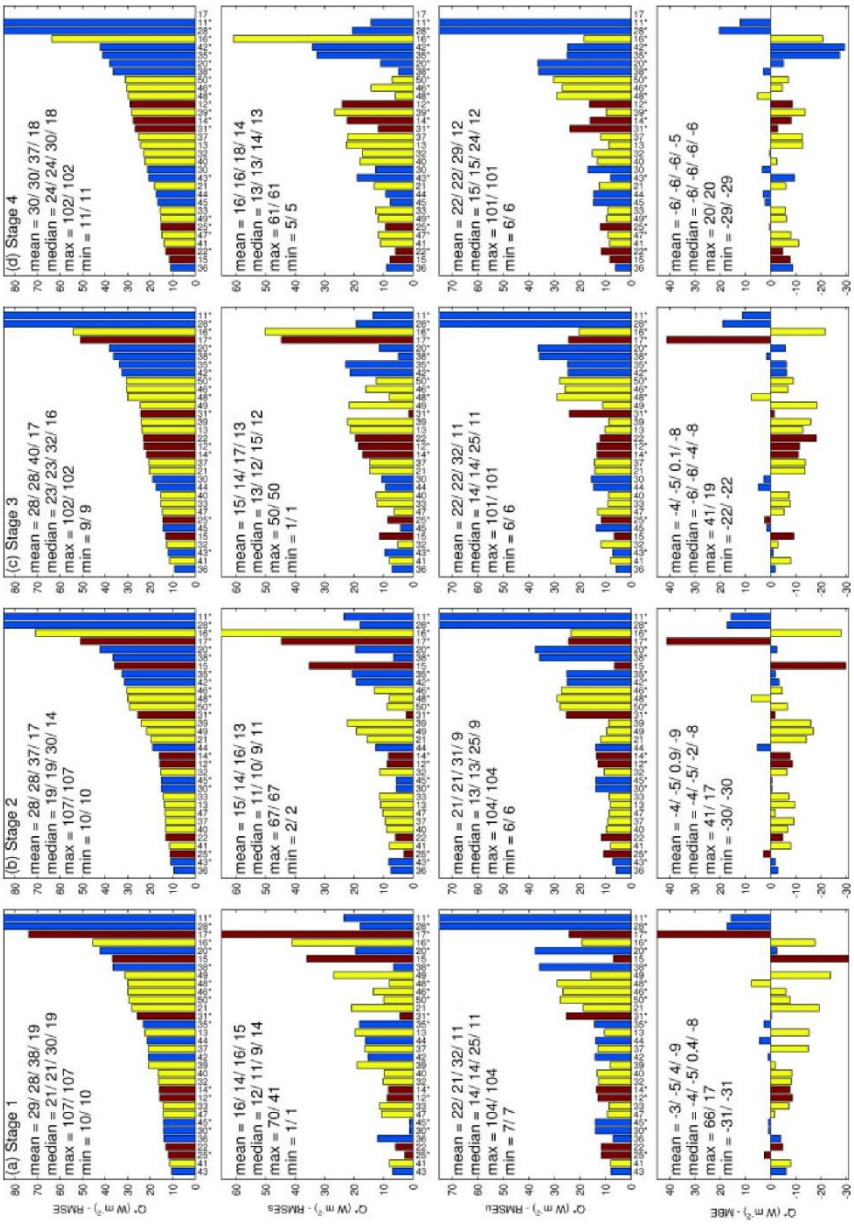

Figure 8. As for Figure 2 but for net all wave radiation (Q^*) for all hours. The mean observed flux for this period was 78.9 W m^{-2}. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

for improvement by those that do include vegetation. The lowest median RMSE at Stage 4 (21 W m^{-2}) is for Vi models, but as for $L_{\uparrow}$, the performance deteriorates from 13 W m^{-2} at Stage 3. The best performing morphology class at Stage 4 is the simplest (L1) but the best performance across all stages and classes is Stage 2 L3, with a median RMSE of 14 W m^{-2}. This is the same result when the models are sorted by their approach to facets and orientation (FO) for the simplest models (FO1) at Stage 2, although FOo is only slightly larger at the same stage. This result is repeated again for classification based on treatment of R1 and for AE1.

The models with radiative closure (P0) have their lowest median RMSE at Stage 2 (15 W m^{-2}) and their largest at Stage 4 (25 W m^{-2}). The smallest median RMSE for P1 models is Stage 3 but these models have the largest IQR in Stages 3 and 4 (Figure 10(e)). As for $L_{\uparrow}$ at Stage 4, the complex (Cc) models perform slightly better than the less complex models even though they have deteriorated from better performance at earlier stages. The Cm models perform least well as a group with an increasing median RMSE with each stage.

The models perform generally better at night than for the 24 h period or for the daytime period (mean observed flux day = 216.83, night = −59.45 W m^{-2}). The night-time median RMSE for Stages 1–4 are 11/10/10/12 W m^{-2} and the median MBE are −7/−7/−2/1 W m^{-2}. At Stage 4, the best performing (median RMSE W m^{-2}) models have Vs (11)/L1&L2 (10)/FO1 (7)/R1 (7)/AE1 (7)/Cs (9) characteristics. The daytime performance for Stages 1–4 for the median RMSE was 27/24/28/29 W m^{-2} and for the median MBE was −5/−5/−8/−12 W m^{-2}. At Stage 4, the best performing (median RMSE W m^{-2}) models have Vi (28)/L1 (25)/FO1 (21)/R1 (25)/AE1 (21)/Cc (27) and Cs (28) characteristics. Compared to $L_{\uparrow}$ there is much greater variability between classes; e.g. the Cm models have daytime median RMSE of 50 W m^{-2} at Stage 4.

Models defined by simpler characteristics often perform best driven by the treatment of solar radiation. However, accounting for vegetation is important in improving the performance of the models. But when the overall complexity of the model is considered it is the more complex models that perform best overall and as a cohort make better use of the new site characteristics provided. The medium complexity models systematically drop in performance with increasing information provided, although there is consistently a Cm-type model performing best throughout.

3.2. Turbulent sensible heat flux

Model errors are larger for the turbulent sensible heat flux (Q_H) than for the radiative fluxes (compare Figures 2, 5, 8, and 11 and Figures 3, 6, 9, and 12). As for the radiative fluxes, the provision of information about the fraction of vegetation (Table II) results in an improvement with a reduction in median RMSE from 62 to 55 W m^{-2} (32 models). A similar sized reduction, down to 49 W m^{-2}, is evident at Stage 3, but at Stage 4 there is a small deterioration in performance (51 W m^{-2}). Throughout, the $RMSE_S$ is smaller than the $RMSE_U$, suggesting that overall RMSE is substantially driven by variability

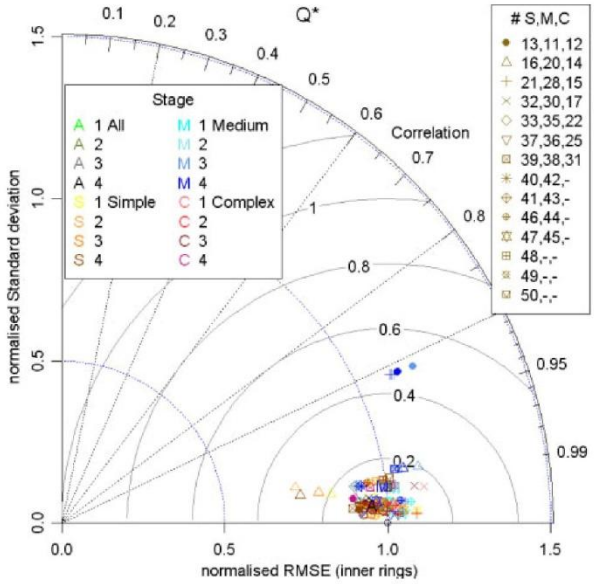

Figure 9. As for Figure 3 but for net all wave radiation (Q^*) for all hours. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

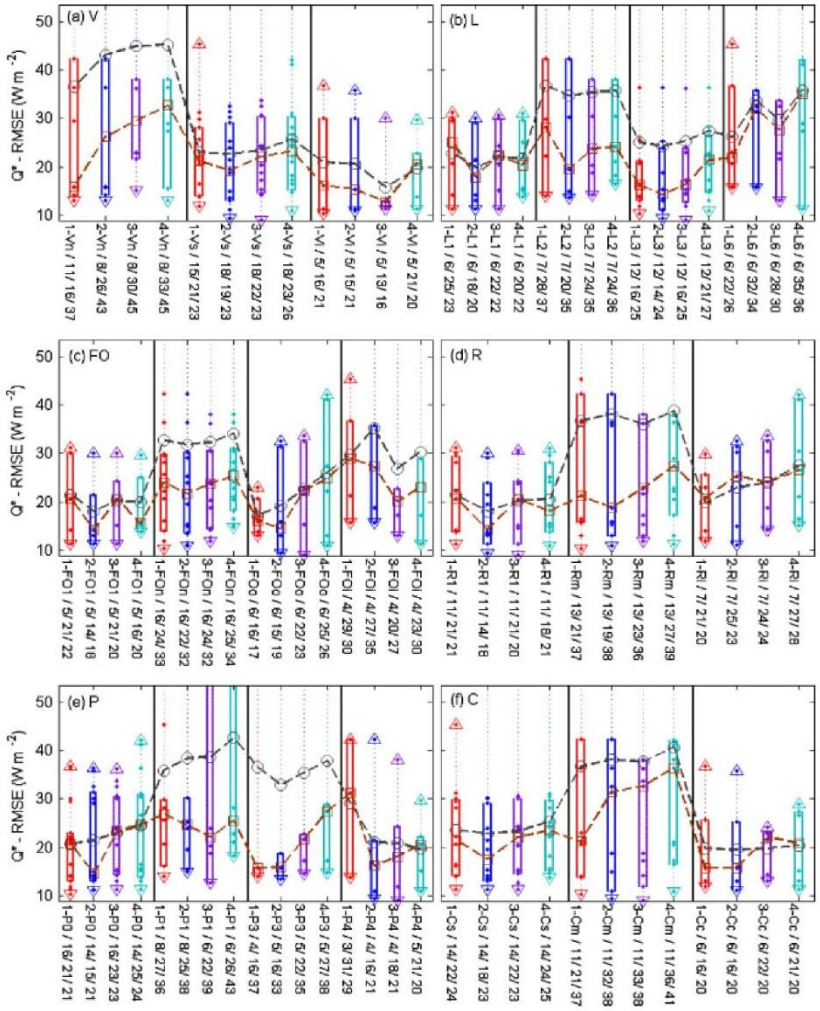

Figure 10. As for Figure 4 but for net all wave radiation (Q^*) for all hours. Note plots are cut-off at 0.50 of the maximum. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

in the observed processes not included in the model physics and is less subject to improvement by better parameter specification. These may be at time scales the models do not capture. The median $RMSE_S$ drops (36/31/23/22 W m^{-2} – 31 models) at each stage as more information is provided about the site but unsystematic error remains around 42 W m^{-2} from Stage 2. The MBE is positive for most models and remains positive at all stages. The largest change in median MBE is at Stage 2, with a reduction from 20 to 6 W m^{-2}. In Stage 3, it rises slightly and then again at Stage 4. Overall there are five models with reduced RMSE at each stage (18 improved from Stage 1 to 2, 10 of which improved from Stage 2 to 3). There are also models whose performance deteriorates between stages; e.g. seven models from Stage 1 to 2 and of those two have a further increase in RMSE at Stage 3. From Stage 2 to 3, 11 models decline in performance (20 improved) followed by four which continue to increase their RMSE (10 improved) at the next stage. From Stage 3 to 4, 17 models improved (14 declined) in performance.

The model which performs best (or second best at Stage 4) is the model which did best for $K_{\uparrow}$, although it did not do best for Q^* or $L_{\uparrow}$. However, the daytime radiation should be reasonable because the shortwave dominates. The performance does not markedly improve through the stages for this model (i.e. there is not a

large reduction in the RMSE). At Stage 1, there are six models with RMSE which have a step drop in performance relative to the others (>10 W m^{-2}). None of these models have radiative closure. In the four stages the poorest model remains the same and has only a 7 W m^{-2} improvement as additional site data became available. Both the best and worst models in Stage 1 do not significantly improve by Stage 4, indicating that they are not benefiting from additional information. However, there is improvement within the middle range of models, most notably model 16 which performs best in Stage 4. The behaviour of the individual models with respect to systematic error shows some slightly surprising results. For example, model 50 which performs poorly overall has almost the smallest RMSE$_S$ overall. In fact, the small RMSE$_S$ are distributed throughout the range of the RMSE (Figure 11).

The normalized Taylor plots (Figure 12) show that no models or the ensembles have a correlation of 0.96 or greater. The ensemble mean values show generally better performance than the individual models. The ensemble of the simple models is the best with a clear change in performance between stages. When all models are included in the ensemble there is clear improvement from Stage 1 to 2 and 3 but reduced performance at Stage 4 (although it is slightly better than Stage 1). The modelling of Q_H is clearly poorer than the radiative fluxes and much more variable than for the radiative fluxes (e.g. compare Figures 9 and 12).

The models without radiation balance closure problems (P0) have a lower median RMSE than those that do not close (P1, P3, P4), except at Stage 4 (P4) (Figure 13(e)) when there is a rise in the median RMSE. For P1 (models which did not use the provided data) and P4 (unknown explanations), there is a reduction in RMSE across stages. Here we do not consider energy balance closure because the details of how Q_F enters the models are critical. Given the different assumption models made (Figure 1), it appears as an input, internal model assumption, and calculated output. At this stage we do not have all these values.

The impact of how vegetation is considered is seen clearly when comparing the Vn models to the Vs and Vi (Figure 13(a)). The Vn models have the widest range, largest IQR, and the poorest median performance. The Vi models perform the best but have a slight decrease in performance at Stage 4. The Vs cohort has the greatest improvement through the stages but also have a decrease at Stage 4. This suggests more complex and realistic treatments of vegetation may be important for modelling Q_H.

The simplest models with respect to morphology (L1) perform best relative to the others and improve across the stages (Figure 13(b)). The L2 models show the largest change between stages. The models which have a canyon but do not account for facet orientation (FO1) have the smallest median RMSE throughout and a steady reduction in the mean RMSE (Figure 13(c)). The treatment of surface temperature (Figure 1) for the built (B) fraction (Z_B) deteriorates with increasing complexity (not shown). The simplest (Z_B1) had an improvement at each stage with the median RMSE improving from 62 to 39 W m^{-2} across the four stages. In the other two approaches a steady improvement is not seen.

The treatment of AN varies from not including it or assuming it is negligible (ANn), to prescribing a value (ANp), to modelling it explicitly, or to using an internal temperature (ANc combined code of ANi, ANm, Figure 1). The simplest (ANn) has the lowest median RMSE and improves steadily across the four stages. Overall, the simplest models (Cs) have the smallest median RMSE at each stage, with improvements evident at each stage (Figure 13(f)). The median RMSE at Stage 4 for the three approaches with increasing level of complexity are 42/55/73 W m^{-2} (Cs/Cm/Cc). Thus, the simpler models often showed a net improvement with additional information, whereas that was not the case for the more complex models. This may be because there was not enough additional detailed information provided for the more complex models so it was more difficult for the users to decide how to use this information appropriately. In addition, such models typically have many more parameter values that could be altered in response to the new information provided.

The daytime results at Stage 1 have a larger median RMSE than the 24 h or night-time (79/62/28 W m^{-2}) which continues to Stage 4 (68/51/21 W m^{-2}). Obviously, the variability and the magnitude of Q_H is much greater during the daytime than for night-time hours (mean observed flux: day = 88.72, night = −13.16 W m^{-2}). The median daytime MBE is positive during the day (40/25 W m^{-2} Stage 1/4) and negative at night by Stage 4 (10/−8 W m^{-2} Stage 1/4). At night, there is one poor model (17) for the three stages, but there is another model that performs very poorly at Stage 3 but in Stage 4 returns to much better performance. These individual model RMSE results are >115 W m^{-2} compared to under <50 W m^{-2} for the remainder of the models. The poorly performing models during the daytime are different and the same two models perform poorly throughout (the difference to the next models is of the order of 50 W m^{-2}). Thus, the models that are performing least well on the all hour basis are caused by different abilities related to day- and night-time processes.

Overall, the simple complexity (Cs) models perform best but it is important to include vegetation. With additional information the models improve but the simplest models have a systematic improvement at each stage, whereas for the more complex models this is not the case. In this case, where Q_F is not very large, the models that do not account for Q_F do better. The slab or bulk models also show a consistent improvement at each stage.

3.3. Turbulent latent heat flux

The modelling of latent heat flux (Q_E) needs to deal with the loss of water from a wet surface, e.g. after rainfall from roofs, roads, and vegetation; and the transpiration

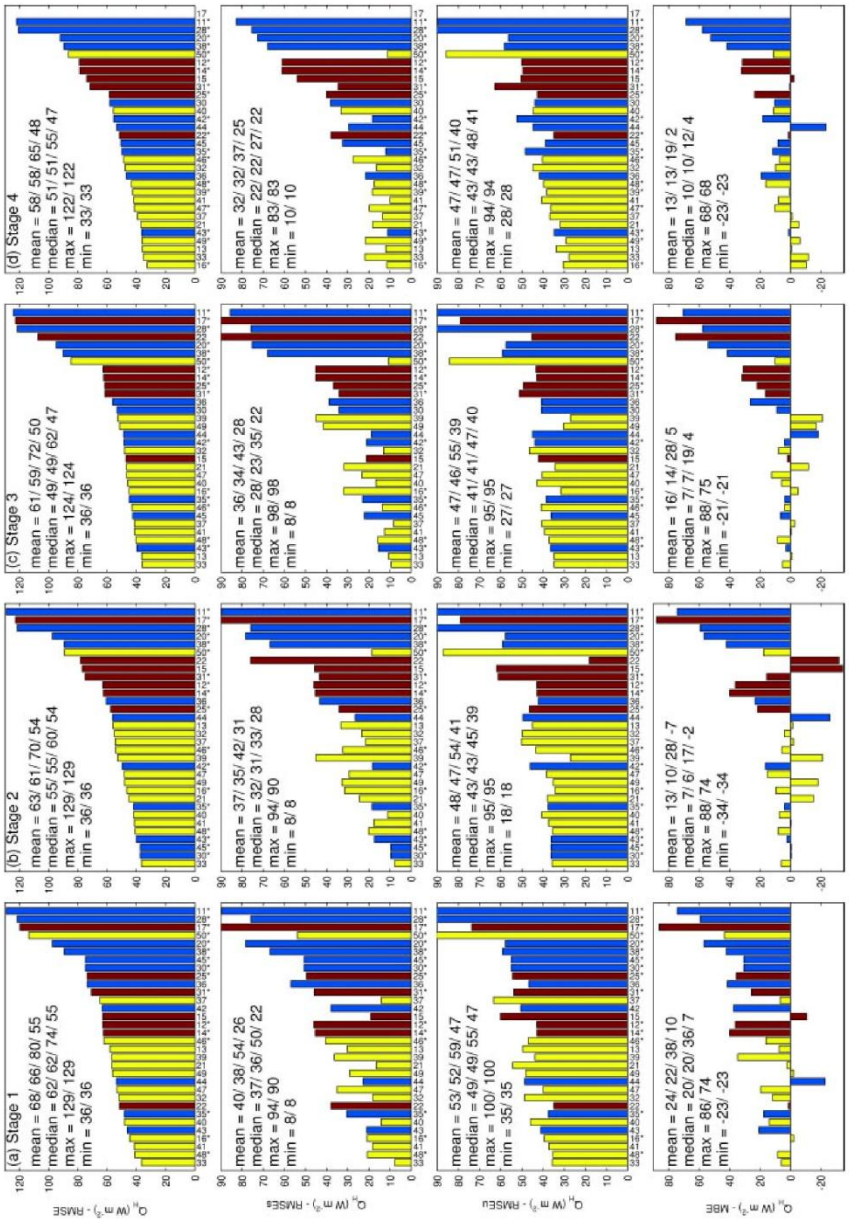

Figure 11. As for Figure 2 but for turbulent sensible heat flux (Q_H) for all hours. The mean observed flux for this period was 37.9 W m^{-2}. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

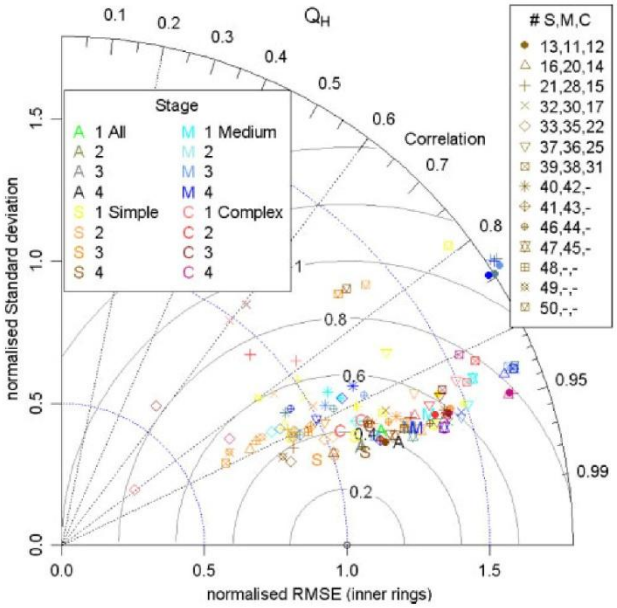

Figure 12. As for Figure 3 but for turbulent sensible heat flux (Q_H) for all hours. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

of vegetation which continues between rainfall events. The median RMSE for the modelled latent heat flux (Figure 14) dropped by the largest amount at Stage 2 when information about the vegetation was provided (54/42/42/43 W m^{-2}, for 31 models). There was no general improvement from knowing more details about the plan area fractions of vegetation (e.g. grass *vs* nongrass, Stage 3). Across the four stages there are six (seven Stage 1) models that have a large $RMSE_S$ (58 W m^{-2}) and a 0 W m^{-2} $RMSE_U$; these are ignoring latent heat flux completely. There are a couple of models that address some aspect of this flux but have even poorer performance than those that neglect it. However, all but one of these models improves so by Stage 4 there is only one model that is in this category. It should be noted that this model does not close the radiation or the energy balance.

From Stage 1 to 2, 17 models have a reduced RMSE; 11 of which improve at Stage 3; and of these, four improve at Stage 4. In the reverse direction, of the eight models which have an increase in RMSE at Stage 2; three have a further increase at Stage 3 and one deteriorates again at Stage 4. Similarly, there is one model that has the largest increase in $RMSE_S$ at Stage 2 and retains this across the stages.

Overall, the systematic errors are generally larger than the unsystematic errors. As noted above, this is largely due to the models not attempting to model latent heat flux (Figure 14). By Stage 4, the median $RMSE_S$ has dropped by nearly 20 W m^{-2}, whereas the $RMSE_U$ remains about the same so there is a definite benefit from the new information provided (either directly as parameters or recognizing the need to consider particular processes more fully). Overall there is a negative MBE, with a median of -18 W m^{-2} at Stage 1. The best performing models based on MBE at Stage 1 have a small positive MBE but the majority have a negative MBE. By Stage 2, the MBE halved to -9 W m^{-2}. This obviously remains large because of those models that have not modelled Q_E but does suggest that those that do include it are generally underestimating the flux. This could be because they do not account for additional urban sources of water through irrigation, which can influence evaporation rates and soil moisture (Grimmond and Oke, 1991). This information was not provided at any of these stages to the model participants.

Initially, except for one Cc model, all the best performing models are simple models and the Cm are all grouped at the poorer performing end (Figure 14). However, at Stage 2, when vegetation fraction became known, Cm models start to improve. By Stage 4, we have all model types represented at the poor end, but the five models with the lowest RMSE are Cs.

The correlation coefficient for all models at all stages is less than 0.8 (Figure 15). This result along with the other normalized statistics on the Taylor plot, demonstrates that Q_E is the least well-modelled flux (compare Figures 3, 6, 9, 12, and 15). There is even wider scatter amongst the models than for Q_H. The ensemble performances generally have the better correlations but the normalized

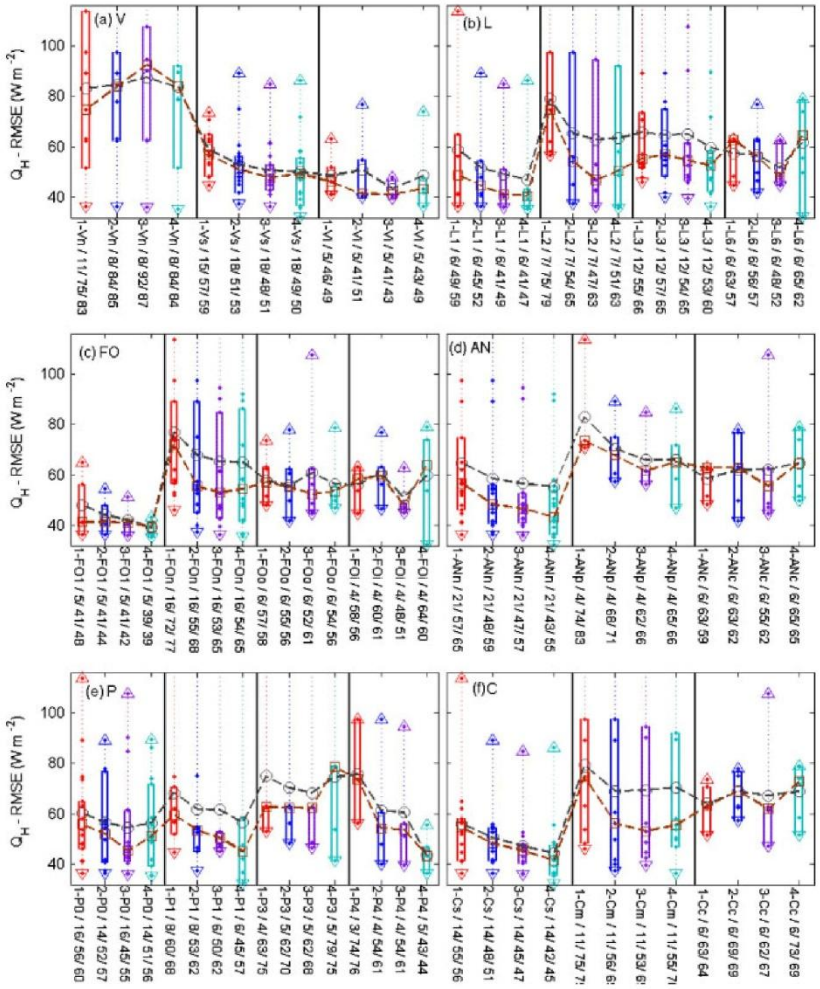

Figure 13. As for Figure 4 but for turbulent sensible heat flux (Q_H) for all hours. Note plots are cut-off at 0.90 of the maximum. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

RMSE are small. The best ensemble performance is for the simple models, followed by the all ensemble. After Stage 1, the medium and complex models have a very similar ensemble performance.

From Stage 1 to 2, three more models chose to include vegetation (Figure 16(a)). The three models which incorporated vegetation did so by using separate vegetation tile(s) (Vs). The Vs approach, the most common, had a 10 W m^{-2} improvement between Stages 1 and 2. This is because in Stage 2 the separate tiles can be more realistically weighted. For Vs models, there is a reduction in the mean RMSE at each stage. For the Vs models, except after rainfall, the latent heat flux is coming exclusively from the vegetation scheme that has been 'coupled' to the urban scheme. These schemes have been extensively tested in earlier PILPS studies; however, they have not been extensively tested for use in urban areas. The user has to decide which vegetation type to select (see discussion in Grimmond *et al.*, 2010) as well as the appropriate parameter values for that vegetation class.

The simpler models which take a bulk approach to the urban morphology (L1) initially have the smallest median RMSE compared to more complex models (L2, L3, L6) (43/58/56/56 W m^{-2}) (Figure 16(b)). The L1 models do improve with subsequent stages but the range also becomes larger. The improvement, however, is not as great by Stage 4 as that which occurs for the L2/L3/L6 (39/38/45/48 W m^{-2}). The L2 models thus improve the

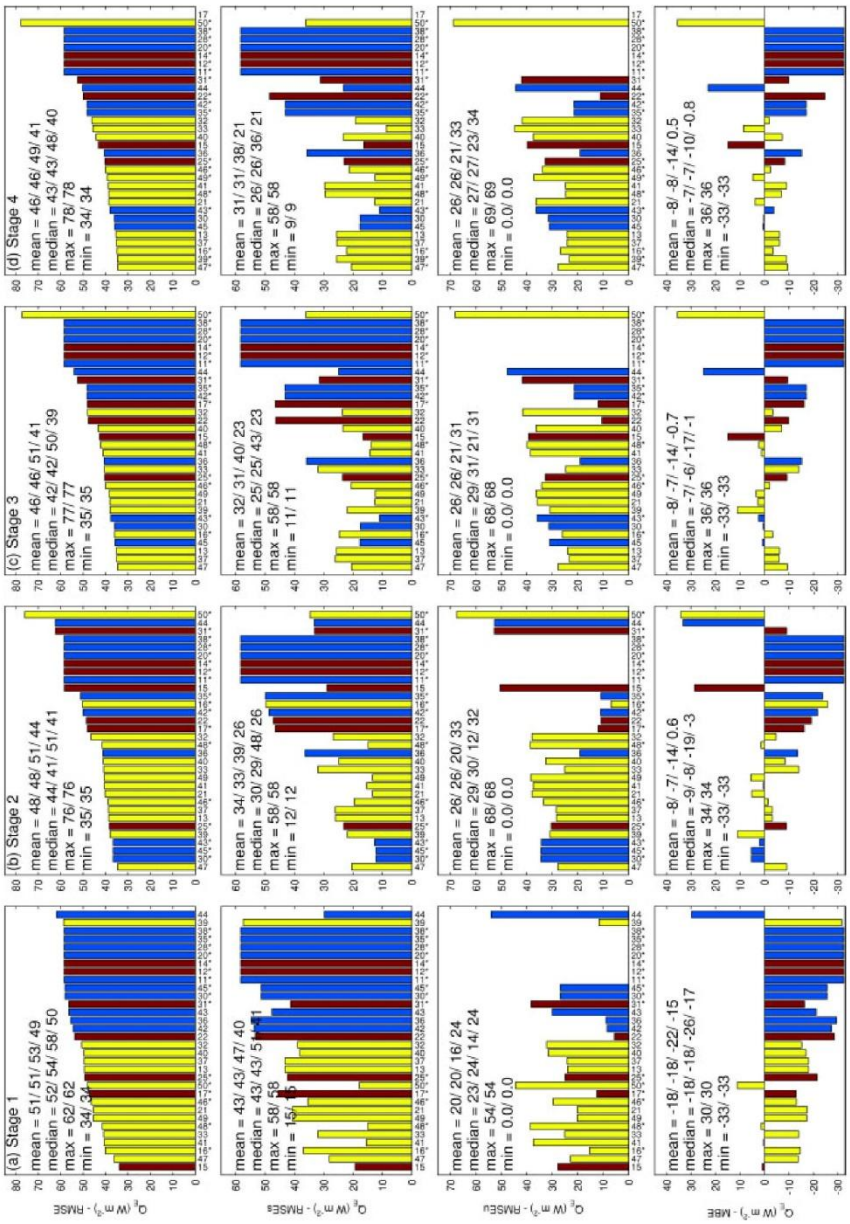

Figure 14. As for Figure 2 but for turbulent latent heat flux (Q_E) for all hours. The mean observed flux for this period was 32.5 W m^{-2}. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

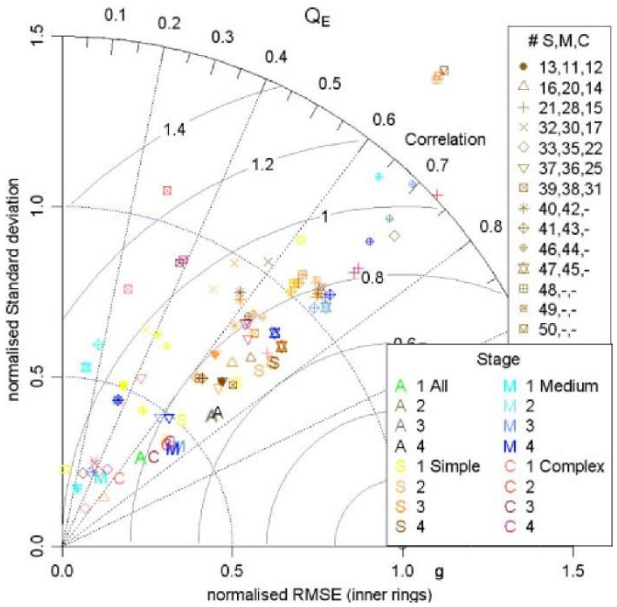

Figure 15. As for Figure 3 but for turbulent latent heat flux (Q_E) for all hours. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

most and have the lowest median RMSE. A small improvement is seen in the median across all four facet and orientations classes (FO) by Stage 4. The models which do not distinguish facets (FO1) have the smallest median RMSE at Stage 4 but the greatest improvement is for those models that have facets but do not account for orientation (FOn).

The models with radiative closure (P0) have a larger median RMSE at Stage 4 than the P1 and P4 models. The P4 models have improvement at each of the four stages but have a slightly larger median RMSE at Stage 4 than the P1 models. The P3 models show no change in the median with stage as many do not model vegetation. Overall, the simplest models (Cs) perform best at all four stages but the Cm models have a greater gain from the additional information provided across the four stages (Figure 16(f)).

The daytime RMSE values are larger than for the night-time period (Stage 1 median 71/21 W m^{-2}) and all hours which is when the observed flux is larger and more variable (mean observed flux day = 56.41 W m^{-2}, night = 8.53 W m^{-2}). The night-time fluxes do not show any improvement in performance over the four stages and there is little variation between methods. At Stage 4, the daytime RMSE is 57 W m^{-2}. The simplest models (Cs) have a median RMSE that is the smallest with a RMSE of 51 W m^{-2} and have a 10 W m^{-2} improvement over the four stages.

The turbulent heat fluxes are not modelled as well as the radiative fluxes. But as with the radiative fluxes the inclusion of vegetation improves model performance. However, despite in Stage 4 knowing the site location, many models did a poorer job than at previous stages.

Overall, the simple models (Cs) do the best job of modelling latent heat flux. They also systematically improve as the additional information becomes available. Taking vegetation into account is critical to model Q_E appropriately. The models that use the separate tile scheme have about the same overall performance as those that take an integrated approach. But there is a much wider range of results from the separate tile models. This suggests that using vegetation schemes that have been tested in nonurban areas are better than ignoring vegetation, but given the wide range of results it suggests that some careful thought may need to be given to ensure their use is appropriate. Here we have not investigated whether the modellers assumed any additional water, such as irrigation, to be available for evaporation.

4. Conclusions

Groups around the world have run ULSMs in offline mode for four stages, with increasing information about the site provided. Initially, the groups knew only that the site was urban but by Stage 4 detailed surface materials characteristics had been provided. Here the ability to model the radiation and energy balance fluxes on average for a year is evaluated. It should be remembered that observations also have errors which vary with time of day, season, latitude, local geography, and land cover.

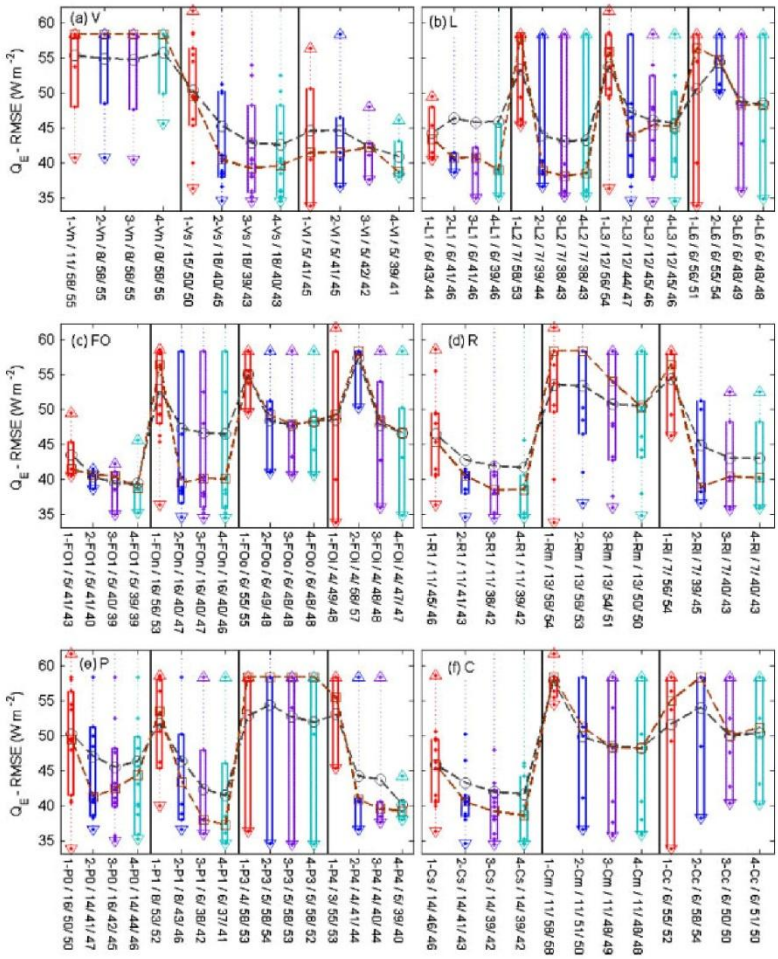

Figure 16. As for Figure 4 but for turbulent latent heat flux (Q_E) for all hours. Note plots are cut-off at 0.80 of the maximum. This figure is available in colour online at wileyonlinelibrary.com/journal/joc

In the process of running the models, a small number of models improved for an individual flux at each stage as new information was provided (2, 2, 2, 5, 4; for $K_\uparrow$, $L_\uparrow$, Q^*, Q_H, Q_E, respectively). However, there are other models that have a drop in performance with the addition of more complete information, and cases where there is a systematic decline at all stages (0, 1, 4, 0, 1; for $K_\uparrow$, $L_\uparrow$, Q^*, Q_H, Q_E, respectively).

From the analysis of the data returned from the modelling groups in relation to the observed flux data the following conclusions are drawn:

- A wide range of model performance is evident for each flux. No individual model does best for every flux modelled. Clearly this finding has very significant implications for the application of any model. It may also imply that in some cases models perform well but for the wrong physical reasons. For example, if a model overestimates the net shortwave radiation, but accurately models the sensible heat flux, then it may indicate a problem also in the physical representation of the heat exchanges between the surfaces and the atmosphere (since it needs to 'absorb' more energy to get the right sensible heat flux).
- Taking vegetation cover into account (or not) significantly impacts model performance. This conclusion is in agreement with those of Phase 1 (Grimmond *et al.*, 2010) where the site had a much lower plan area fraction of vegetation than the Phase 2 site. Data provided at Stage 2 (surface cover fractions) usually had the

largest impact on model performance. Moreover, the fact that the RMSE for the latent heat flux is of the same order as the latent heat flux itself, indicates that work needs to be done to improve simulations of this flux.

- Closure of the radiation balance is not a good measure of the ability to calculate a particular radiative flux. Comparing the performance of the components of the radiation balance to the net all wave radiation shows a clear re-ranking between fluxes. Notably those that perform poorly for an individual component flux are not the poorest for Q^* (Figures 2, 5, and 8). This means it is important that when a user applies a model they are aware of the performance of the ULSM not only for the initial flux of interest but also for the other fluxes for which the user may wish to infer impact. Given the increasing use of ULSM for assessing mitigation and adaptation strategies this is very important.
- Overall, the ULSM generally model $K_{\uparrow}$ well and additional surface information does result in an improvement of performance. The models are able to estimate reasonably well the amount of energy absorbed by the urban fabric, but have bigger problems in partitioning it between longwave, sensible, latent, and storage heat fluxes.
- Overall $L_{\uparrow}$ is not as well modelled as $K_{\uparrow}$. The set of model characteristics that minimize the errors in the outgoing longwave radiation change with the time of day. Generally performance improved when the pervious/impervious fraction became known but did not when heights and further information on surface fractions were provided. The performance of most models deteriorated when building material information was provided; typically back to the levels at Stage 1 but in many cases even poorer. Given the difficulty to gather appropriate values of material characteristics, their provision may not currently be worth the effort given how models then perform. Alternatively, there is a need to ensure that the data are of much better quality than is currently 'easily' obtainable.
- Net all wave radiation is modelled better than either $K_{\uparrow}$ or $L_{\uparrow}$. In general, the radiative fluxes are modelled better than the turbulent fluxes. The net all wave radiation is clearly the best modelled flux which is in agreement with Phase 1 results (Grimmond *et al.*, 2010). There is clear trade-off in performance between net all wave radiation (Q^*) and turbulent sensible heat flux (Q_H) which is in agreement with Loridan *et al.* (2010a).
- The errors from the models were smaller during the night than they were during the daytime, although this might be expected as the surface energy balance is not dominated by the solar radiation during this period.
- For the net radiation, simple characteristics (L1/FO1/R1/AE1/Cs) give the best results for both daytime and night-time, although there is much greater variability between the classes than for the outgoing longwave radiation.
- The models that perform best, for individual characteristics, are those that are the simplest as they can be assigned one parameter that is close to the observed value. Based on overall complexity the simplest and the most complex have similar results which are better than the medium complexity models.
- Additional surface information is important in improving model performance. However, there is evidence that good model physics is not enough to prevent the users' choice of parameter values from significantly influencing the outcome. Therefore, it is essential when models are being used for scenario testing that appropriate parameter values are used.
- Simpler models often showed a net improvement with additional information; the more complex models did not. This may be because there was not enough additional detailed information provided so it was more difficult for the users to decide how to use this information appropriately. It is important to note that parameters specified for simpler models (e.g. overall albedo) often equate to empirical aggregations of processes in more complex models (e.g. the net effect of reflections due to facet albedos). Nevertheless, the results here suggest that increased model complexity does not necessarily increase model performance.
- It is expected that more complex models may have more potential for future improvements as they are able to resolve more details without deteriorating their performance. The most complex models are more flexible and have the potential to describe the biophysical interactions between the atmosphere and urban surfaces. Although the ability to do this has not been tested here, these models can provide vertical profiles of atmospheric variables within the urban canopy layer. If the simulation is for weather forecasting, a good estimate of the heat fluxes at the top of the urban canopy is probably sufficient, and, consequently, a simple scheme may be the appropriate choice. If air quality is the focus, the atmospheric behaviour within the urban canopy layer may be important, and a more complex scheme can be useful. An important finding of PILPS-urban is that in many cases, work is needed to improve the complex schemes (both in terms of physics and definition of numerical constants), in order to have skills comparable to those of the more simple schemes in estimating energy fluxes at the top of the urban canopy. More complicated models are generally more difficult to use and it is even difficult for modellers to identify which are the most critical points of their model.
- As a community it is clear that in terms of surface characteristics, the information up to Stage 3 (Table II) benefited a large number of models. The AE were also beneficial (Stage 4) but the provision and acquisition of the most appropriate wall, roof, and road thermal properties need further thought and development from the modelling community. This model intercomparison

has already generated a suggested improved method for thermal parameter specification that accounts for the high degree of heterogeneity of such parameters in cities (Salamanca *et al*. 2009, 2010). Hopefully, additional analyses will shed more light on this issue.
- Using an ensemble of models rather than one model is generally better than any individual model for an individual flux. In general, the medium complexity ensemble performs least well and the simple performs best. The 'all' ensemble is always better than the medium complexity. Given the overall better performance of the ensembles they may be better than using one individual model when considering all of the fluxes.

These results are the first of a number of different studies that will be undertaken from these model runs. Future analyses will consider the role of seasonality on model performance, role of cloud conditions (day and night), time since rainfall, wind regime, the range of parameter values that are used, and the determination of optimized parameters; the participants will also analyse what they have learnt from the model comparison. To date, only two urban sites have been compared (Phase 1 and 2), which obviously is not representative of the wide range of land covers and morphologies, etc. found within neighbourhoods around the globe. However, some common conclusions are arrived at from comparison with these two sites, such as the best ability is for modelling net all wave radiation flux. Most notably, despite the range of vegetation cover found at the two sites, accounting for vegetation appears to be essential when modelling urban surface energy flux exchanges. There is a need for future comparisons of this type for sites with varying morphology and across a wider range of building materials. Our initial message is one of caution in applying any ULSM because, in general, no model performs well across all fluxes and it may be best to use an ensemble approach.

Acknowledgements

We would like to thank the IAUC/ICUC 7 editors of this volume. Funds to support this work (Grimmond) have included: Met Office (P001550), EU framework 7 (7 FP7-ENV-2007-1) projects MEGAPOLI (212520), BRIDGE (211345), and NSF (ATM-0710631). This work contributes to COST728. We would like to thank all who were involved in the collection of the original dataset and organisations and individuals who provided data that supported the work. The following people provided information about the site characteristics: Faculty of Architecture, Building and Planning, University of Melbourne (David O'Brian, John Sadar, Hamish Hill, Jon Robinson, Anna Hurlimann, Julie Willis, Valerie Francis, Carolyn Whitzman), Centre for Sustainable Architecture with Wood, School of Architecture & Design, University of Tasmania (Gregory Nolan), City of Darebin Council (David Archer), TPC Solutions Pty. Ltd. (Boris Iskra), Forest and Wood Products Australia, Ltd. (Jugo Ilic), Centre for Sustainable Infrastructure, Civil Engineering, Swinburne University of Technology (Kerry McManus). Funding from CATER 2006-2202 (Baik), BSIK-COM29 (Steeneveld), and all the agencies that support the considerable time contributed by participating groups are acknowledged.

References

Best MJ. 2005. Representing urban areas within operational numerical weather prediction models. *Boundary-Layer Meteorology* **114**: 91–109.

Best MJ, Grimmond CSB, Villani MG. 2006. Evaluation of the urban tile in MOSES using surface energy balance observations. *Boundary-Layer Meteorology* **118**: 503–525.

Chen F, Kusaka H, Tewari M, Bao J, Hirakuchi H. 2004. Utilizing the coupled WRF/LSM/Urban modeling system with detailed urban classification to simulate the urban heat island phenomena over the Greater Houston area. *Fifth Symposium on the Urban Environment*, CD-ROM. 9.11. Amer. Meteor. Soc., Vancouver, BC, Canada.

Clarke JA, Yaneske PP, Pinney AA. 1991. *The Harmonisation of Thermal Properties of Building Materials*. BEPAC Publication: Watford, UK; TN91/6, ISBN 0 187 212 607 3.

Coutts AM, Beringer J, Tapper NJ. 2007a. Characteristics influencing the variability of urban CO2 fluxes in Melbourne, Australia. *Atmospheric Environment* **41**: 51–62.

Coutts AM, Beringer J, Tapper NJ. 2007b. Impact of increasing urban density on local climate: spatial and temporal variations in the surface energy balance in Melbourne, Australia. *Journal of Applied Meteorology* **47**: 477–493.

Dandou A, Tombrou M, Akylas E, Soulakellis N, Bossioli E. 2005. Development and evaluation of an urban parameterization scheme in the Penn State/NCAR Mesoscale model (MM5). *Journal of Geophysical Research* **110**: D10102.

Dragoni D, Schmid HP, Grimmond CSB, Loescher H. 2007. Uncertainty of annual net ecosystem productivity estimated using eddy-covariance flux measurements. *Journal of Geophysical Research* **112**: D17102, DOI:10.1029/2006JD008149.

Dupont S, Mestayer PG. 2006. Parameterisation of the urban energy budget with the submesoscale soil model. *Journal of Applied Meteorology and Climatology* **45**: 1744–1765.

Dupont S, Mestayer PG, Guilloteau E, Berthier E, Andrieu H. 2006. Parameterisation of the urban water budget with the submesoscale soil model. *Journal of Applied Meteorology and Climatology* **45**: 624–648.

Engineering Toolbox. 2005a. Gases – specific heat capacities and individual gas constants [Online]. Available from: http://www.engineeringtoolbox.com/spesific-heat-capacity-gases-d_159.html (Last accessed March 2010).

Engineering Toolbox. 2005a. Thermal conductivity of some common materials [Online]. Available from: http://www.engineeringtoolbox.com/thermal-conductivity-d_429.html (Last accessed March 2010).

Essery RLH, Best MJ, Betts RA, Cox PM, Taylor CM. 2003. Explicit representation of subgrid heterogeneity in a GCM land surface scheme. *Journal of Hydrometeorology* **4**: 530–543.

Foken T. 2008. *Micrometeorology*. Springer: Berlin, Heidelberg; 308 pp, ISBN: 978-3-540-74665-2.

Fortuniak K. 2003. A slab surface energy balance model (SUEB) and its application to the study on the role of roughness length in forming an urban heat island. *Acta Universitatis Wratislaviensis* **2542**: 368–377.

Fortuniak K, Offerle B, Grimmond CSB. 2004. Slab surface energy balance scheme and its application to parameterisation of the energy fluxes on urban areas. NATO ASI, Kiev, Ukraine; 82–83. Available from: www.met.rdg.ac.uk/urb_met/NATO_ASI/talks.html (Last accessed 4–15 May 2010).

Fortuniak K, Offerle B, Grimmond CSB. 2005. Application of a slab surface energy balance model to determine surface parameters for urban areas. *Lund Electronic Reports in Physical Geography* **5**: 90–91.

Gillett NP, Zwiers FW, Weaver AJ, Hegerl GC, Allen MR, Stott PA. 2002. Detecting anthropogenic influence with a multi-model ensemble. *Geophysical Research Letters* **29**: 1970, DOI:10.1029/2002GL015836.

Grimmond CSB, Oke TR. 1991. An evaporation-interception model for urban areas. *Water Resources Research* **27**: 1739–1755.

Grimmond CSB, Oke TR. 1999. Heat storage in urban areas: observations and evaluation of a simple model. *Journal of Applied Meteorology* **38**: 922–940.

Grimmond CSB, Oke TR. 2002. Turbulent heat fluxes in urban areas: observations and local-scale urban meteorological parameterization scheme (LUMPS). *Journal of Applied Meteorology* **41**: 792–810.

Grimmond CSB, Best M, Barlow J, Arnfield AJ, Baik J-J, Belcher S, Bruse M, Calmet I, Chen F, Clark P, Dandou A, Erell E, Fortuniak K, Hamdi R, Kanda M, Kawai T, Kondo H, Krayenhoff S, Lee S-H, Limor S-B, Martilli A, Masson V, Miao S, Mills G, Moriwaki R, Oleson K, Porson A, Sievers U, Tombrou M, Voogt J, Williamson T. 2009. Urban surface energy balance models: model characteristics and methodology for a comparison study. In *Meteorological and Air Quality Models for Urban Areas*, Baklanov A, Grimmond CSB, Mahura A, Athanassiadou M (eds). Springer-Verlag: Berlin, Heidelberg; ISBN: 978-3-642-00297-7.

Grimmond CSB, Blackett M, Best MJ, Barlow J, Baik J-J, Belcher SE, Bohnenstengel SI, Calmet I, Chen F, Dandou A, Fortuniak K, Gouvea ML, Hamdi R, Hendry M, Kawai T, Kawamoto Y, Kondo H, Krayenhoff ES, Lee S-H, Loridan T, Martilli A, Masson V, Miao S, Oleson K, Pigeon G, Porson A, Ryu Y-H, Salamanca F, Shashua-Bar L, Steeneveld G-J, Trombou M, Voogt J, Young D, Zhang N. 2010. The international urban energy balance models comparison project: first results from phase 1. *Journal of Applied Meteorology and Climatology* **49**: 1268–1292, DOI: 10.1175/2010JAMC2354.1.

Hamdi R, Schayes G. 2007. Validation of Martilli's urban boundary layer scheme with measurements from two mid-latitude European cities. *Atmospheric Chemistry and Physics* **7**: 4513–4526.

Hamdi R, Masson V. 2008. Inclusion of a drag approach in the Town Energy Balance (TEB) scheme: offline 1-D evaluation in a street canyon. *Journal of Applied Meteorology and Climatology* **47**: 2627–2644.

Harman IN, Best MJ, Belcher SE. 2004a. Radiative exchange in an urban street canyon. *Boundary-Layer Meteorology* **110**: 301–316.

Harman IN, Barlow JF, Belcher SE. 2004b. Scalar fluxes from urban street canyons. Part II: model. *Boundary-Layer Meteorology* **113**: 387–410.

Harman IN, Belcher SE. 2006. The surface energy balance and boundary layer over urban street canyons. *Quarterly Journal of the Royal Meteorological Society* **132**: 2749–2768.

Henderson-Sellers A, Yang ZL, Dickenson RE. 1993. The project for intercomparison of land-surface parameterization schemes. *Bulletin of the American Meteorological Society* **74**: 1335–1349.

Henderson-Sellers A, Irannejad P, McGuffie K, Pitman A. 2003. Predicting land-surface climates-better skill or moving targets?. *Geophysical Research Letters* **30**: 1777, DOI:10.1029/2003GL017387.

Hollinger DY, Richardson AD. 2005. Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology* **25**: 873–885.

Irranejad P, Henderson-Sellers A, Sharmeen S. 2003. Importance of land-surface parameterization for latent heat simulation in global atmospheric models. *Geophysical Research Letters* **30**: 1904, DOI:10.1029/2003/GL018044.

Jacobson MZ. 1999. *Fundamentals of Atmospheric Modeling*. Cambridge University Press: Cambridge.

Kanda M, Kawai T, Kanega M, Moriwaki R, Narita K, Hagishima A. 2005a. A simple energy balance model for regular building arrays. *Boundary-Layer Meteorology* **116**: 423–443.

Kanda M, Kawai T, Nakagawa K. 2005b. A simple theoretical radiation scheme for regular building arrays. *Boundary-Layer Meteorology* **114**: 71–90.

Kawai T, Kanda M, Narita K, Hagishima A. 2007. Validation of a numerical model for urban energy-exchange using outdoor scale-model measurements. *International Journal of Climatology* **27**: 1931–1942.

Kawai T, Ridwan MK, Kanda M. 2009. Evaluation of the simple urban energy balance model using 1-yr flux observations at two cities. *Journal of Applied Meteorology and Climatology* **48**: 693–715.

Kawamoto Y, Ooka R. 2006. Analysis of the radiation field at pedestrian level using a meso-scale meteorological model incorporating the urban canopy model. In *ICUC-6*, Göteborg, Sweden, 12–16 June 2006.

Kawamoto Y, Ooka R. 2009a. Accuracy validation of urban climate analysis model using MM5 incorporating a multi-layer urban canopy model. In *ICUC-7*, Yokohama, Japan, 28 June–3 July 2009.

Kawamoto Y, Ooka R. 2009b. Development of urban climate analysis model using MM5 Part 2 – incorporating an urban canopy model to represent the effect of buildings. *Journal of Environmental Engineering (Transactions of AIJ)* **74**(642): 1009–1018 (in Japanese).

Kondo H, Liu FH. 1998. A study on the urban thermal environment obtained through a one-dimensional urban canopy model. *Journal of Japan Society for Atmospheric Environment* **33**: 179–192 (in Japanese).

Kondo H, Genchi Y, Kikegawa Y, Ohashi Y, Yoshikado H, Komiyama H. 2005. Development of a multi-layer urban canopy model for the analysis of energy consumption in a big city: structure of the urban canopy model and its basic performance. *Boundary-Layer Meteorology* **116**: 395–421.

Krayenhoff ES, Voogt JA. 2007. A microscale three-dimensional urban energy balance model for studying surface temperatures. *Boundary-Layer Meteorology* **123**: 433–461.

Kusaka H, Kondo H, Kikegawa Y, Kimura F. 2001. A simple single-layer urban canopy model for atmospheric models: comparison with multi-layer and slab models. *Boundary-Layer Meteorology* **101**: 329–358.

Lee S-H, Park S-U. 2008. A vegetated urban canopy model for meteorological and environmental modelling. *Boundary-Layer Meteorology* **126**: 73–102.

Lee X, Massman WJ, Law BE. 2004. *Handbook of Micrometeorology*, Kluwer Academic Publishers: Dordrecht.

Lemonsu A, Grimmond CSB, Masson V. 2004. Modelling the surface energy balance of an old Mediterranean city core. *Journal of Applied Meteorology* **43**: 312–327.

Loridan T, Grimmond CSB, Grossman-Clarke S, Chen F, Tewari M, Manning K, Martilli A, Kusaka H, Best M. 2010a. Trade-offs and responsiveness of the single-layer urban canopy parameterization in WRF: an offline evaluation using the MOSCEM optimization algorithm and field observations. *Quarterly Journal of the Royal Meteorological Society* **136**: 997–1019, DOI:10.1002/qj.614.

Loridan T, Grimmond CSB, Offerle BD, Young DT, Smith T, Jarvi L. Lindberg 2010 Local-Scale Urban Meteorological Parameterization Scheme (LUMPS): longwave radiation parameterization & seasonality related developments *Journal of Applied Meteorology & Climatology*, Doi: 10.1175/2010JAMC2474.1.

Martilli A, Clappier A, Rotach MW. 2002. An urban surface exchange parameterisation for mesoscale models. *Boundary Layer Meteorology* **104**: 261–304.

Masson V. 2000. A physically-based scheme for the urban energy budget in atmospheric models. *Boundary-Layer Meteorology* **41**: 1011–1026.

Masson V, Grimmond CSB, Oke TR. 2002. Evaluation of the Town Energy Balance (TEB) scheme with direct measurements from dry districts in two cities. *Journal of Applied Meteorology* **41**: 1011–1026.

Masson V, Seity Y. 2009. Including atmospheric layers in vegetation and urban offline surface schemes. *Journal of Applied Meteorology and Climatology* **48**: 1377–1397.

Ochsner TE, Horton R, Renb T. 2001. A new perspective on soil thermal properties. *Soil Science Society of America Journal* **65**: 1641–1647.

Offerle B, Grimmond CSB, Oke TR. 2003. Parameterization of net all-wave radiation for urban areas. *Journal of Applied Meteorology* **42**: 1157–1173.

Offerle B, Grimmond CSB, Fortuniak K. 2005. Heat storage and anthropogenic heat flux in relation to the energy balance of a central European city center. *International Journal of Climatology* **25**: 1405–1419.

Oleson KW, Bonan GB, Feddema J, Vertenstein M, Grimmond CSB. 2008a. An urban parameterization for a global climate model: 1. Formulation and evaluation for two cities. *Journal of Applied Meteorology and Climatology* **47**: 1038–1060.

Oleson KW, Bonan GB, Feddema J, Vertenstein M. 2008b. An urban parameterization for a global climate model: 2. Sensitivity to input parameters and the simulated heat island in offline simulations. *Journal of Applied Meteorology and Climatology* **47**: 1061–1076.

Pigeon G, Moscicki MA, Voogt JA, Masson V. 2008. Simulation of fall and winter surface energy balance over a dense urban area using the TEB scheme. *Meteorology and Atmospheric Physics* **102**: 159–171.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE. 2010. Implementation of a new urban energy budget scheme in the MetUM. Part I: description and idealized simulations. *Quarterly Journal of the Royal Meteorological Society*, DOI:10. 1002/qj.668.

Porson A, Harman IN, Bohnenstengel SI, Belcher SE. 2009. How many facets are needed to represent the surface energy balance of an urban area? *Boundary-Layer Meteorology* **132**: 107–128.

Richardson AD, Hollinger DY, Burba GG, Davis K,J, Flanagan LB, Katul GG, Munger JW, Ricciuto DM, Stoy PC, Suyker AE, Verma SB, Wofsy SC. 2006. A multi-site analysis of random error in tower-based measurements of carbon and energy fluxes. *Agricultural and Forest Meteorology* **136**: 1–18, DOI:10.1016/j.agrformet.2006.01.007.

Roberts SM, Oke TR, Grimmond CSB, Voogt JA. 2006. Comparison of four methods to estimate urban heat storage. *Journal of Applied Meteorology and Climatology* **45**: 1766–1781.

Ryu Y-H, Baik J-J, Lee S-H. 2009. A new single-layer urban canopy model for use in mesoscale atmospheric models. In *Proceedings of the Seventh International Conference on Urban Climate, ICUC-7*, Yokohama, Japan, June 28–July 3 2009.

Sailor DJ, Lu L. 2004. A top-down methodology for developing diurnal and seasonal anthropogenic heating profiles for urban areas. *Atmospheric Environment* **38**: 2737–2748.

Salamanca F, Krayenhoff ES, Martilli A. 2009. On the derivation of material thermal properties representative of heterogeneous urban neighbourhoods. *Journal of Applied Meteorology and Climatology* **48**: 1725–1732.

Salamanca F, Krpo A, Martilli A, Clappier A. 2010. A new building energy model coupled with an urban canopy parameterization for urban climate simulations – part I. Formulation, verification, and sensitivity analysis of the model. *Theoretical and Applied Climatology*, DOI: 10.1007/s00704-009-0142-9.

Salamanca F, Martilli A. 2010. A new Building Energy Model coupled with an Urban Canopy Parameterization for urban climate simulations – part II. Validation with one dimension off-line simulations. *Theoretical and Applied Climatology* **99**: 345–356.

Taylor KE. 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* **106**: 7183–7192.

Willmott CJ. 1981. On the validation of models. *Physical Geography* **2**: 184–194.

Appendix C: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes

Geosci. Model Dev., 4, 677–699, 2011
www.geosci-model-dev.net/4/677/2011/
doi:10.5194/gmd-4-677-2011

Geoscientific Model Development

The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes

M. J. Best[1], M. Pryor[2], D. B. Clark[3], G. G. Rooney[1], R .L. H. Essery[4], C. B. Ménard[4], J. M. Edwards[1], M. A. Hendry[1], A. Porson[1], N. Gedney[2], L. M. Mercado[3], S. Sitch[5], E. Blyth[3], O. Boucher[1,*], P. M. Cox[6], C. S. B. Grimmond[7], and R. J. Harding[3]

[1]Met Office, FitzRoy Road, Exeter, EX1 3PB, UK
[2]Met Office, JCHMR, Wallingford, UK
[3]Centre for Ecology and Hydrology, Wallingford, UK
[4]University of Edinburgh, Edinburgh, UK
[5]University of Leeds, Leeds, UK
[6]University of Exeter, Exeter, UK
[7]King's College London, London, UK
[*]now at: Laboratoire de Météorologie Dynamique, IPSL, CNRS/UPMC, Paris, France

Received: 1 March 2011 – Published in Geosci. Model Dev. Discuss.: 24 March 2011
Revised: 20 July 2011 – Accepted: 22 July 2011 – Published: 1 September 2011

Abstract. This manuscript describes the energy and water components of a new community land surface model called the Joint UK Land Environment Simulator (JULES). This is developed from the Met Office Surface Exchange Scheme (MOSES). It can be used as a stand alone land surface model driven by observed forcing data, or coupled to an atmospheric global circulation model. The JULES model has been coupled to the Met Office Unified Model (UM) and as such provides a unique opportunity for the research community to contribute their research to improve both world-leading operational weather forecasting and climate change prediction systems. In addition JULES, and its forerunner MOSES, have been the basis for a number of very high-profile papers concerning the land-surface and climate over the last decade. JULES has a modular structure aligned to physical processes, providing the basis for a flexible modelling platform.

1 Introduction

Traditionally Land Surface Models (LSMs) have been considered as the lower boundary condition for Global Circulation Models (GCMs) and other atmospheric modelling systems. Over the last couple of decades, the importance of the influence that the land surface has on atmospheric modelling has increased, which has led to additional focus on the complexity and accuracy of LSMs. Models have developed from a simple energy balance with a simple soil scheme (e.g., Deardorff, 1978) through to complex vegetation structures with multiple layer soil hydrology. Examples of currently used land surface schemes include the Interaction Soil-Biosphere-Atmosphere model (ISBA, Noilhan and Planton, 1989), the Canadian Land Surface Scheme (CLASS, Verseghy, 1991; Verseghy et al., 1993), the Tiled ECMWF Scheme for Surface Exchanges over Land model (TESSEL, Viterbo and Beljaars, 1995), the NOAH model (Ek et al., 2003) and the Community Land Model (CLM, Oleson et al., 2010).

The large differences in the response of the surface fluxes to various surfaces has initiated a representation of subgridscale heterogeneity, such as tile or mosaic schemes (e.g., Essery et al., 2003a). Differences at the surface can be caused by their interaction with snow (e.g., snow on top of the surface as with bare soil and short vegetation, or snow under the "surface" as with needleleaf forests), the availability of water at the surface influencing the Bowen ratio (e.g., open water, snow and ice surfaces compared to vegetation and bare soil surfaces), or in the treatment of the carbon cycle for vegetation (e.g., the difference in carbon pathways between C3 and C4 vegetation). Further increases in model resolution, particularly for regional scale operational weather forecasting, open up new challenges in the way we represent the subgridscale heterogeneity at the surface, as the nature of the heterogeneity changes.

Correspondence to: M. Best
(martin.best@metoffice.gov.uk)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Table 1. Meteorological forcing data required to drive the JULES model.

Data	Units
Downward component of shortwave radiation at the surface	$W\,m^{-2}$
Downward component of longwave radiation at the surface	$W\,m^{-2}$
Rainfall	$kg\,m^{-2}\,s^{-1}$
Snowfall	$kg\,m^{-2}\,s^{-1}$
U component of wind	$m\,s^{-1}$
V component of wind	$m\,s^{-1}$
Atmospheric temperature	K
Atmospheric specific humidity	$kg\,kg^{-1}$
Surface Pressure	Pa

As the resolution and accuracy of atmospheric modelling systems increases, there is likely to be a need for a wider diversity of land surface processes, such as river flow and flooding, groundwater, or potential crop yields. These new processes present some challenges as model developers will have to acquire new areas of expertise and integrate new science in existing modelling systems.

The development in our understanding of the interactions between the atmosphere and the biosphere for the carbon cycle has begun a new era for science in land surface modelling (e.g., Cox et al., 2000). Current research activities are not limited to the carbon cycle, but are also considering other elements such as the nitrogen cycle, methane and ozone (Gedney et al., 2004; Sitch et al., 2007; Thornton et al., 2007, 2009; Sokolov et al., 2008; Fisher et al., 2010; Zaehle et al., 2010). Again, the complexity of these new systems require additional expert knowledge that has traditionally not been held by the original LSM developers.

It is beyond most research and operational centres to have the expertise in such a diverse range of science. Therefore to develop a state of the art LSM requires an alternative perspective to the traditional isolated development of these modelling systems. The development of a community land surface model enables experts in areas of land surface science to contribute towards a leading land surface model, from which all users will benefit. This approach has been adopted with the Community Land Model (CLM) and the NOAH model, and now with the new community land surface model, the Joint UK Land Environment Simulator (JULES). JULES originated from the Met Office Surface Exchange Scheme (MOSES; Cox et al., 1999; Essery et al., 2003a), the land surface model developed at the UK Met Office for applications ranging from operational weather forecasting to Earth system modelling. The forcing data required by JULES (Table 1) are the standard information that would be exchanged when coupled to an atmospheric GCM. Hence, JULES can be linked to the UK Met Office Unified Model (Cullen, 1993) opening up the unique opportunity for the research community to contribute its science into leading operational weather forecasting and climate change prediction systems. In addition JULES, and its forerunner MOSES, have already been the basis of a number of high-profile papers on the response of land ecosystems to climate (Cox et al., 2000; Gedney et al., 2006; Betts et al., 2007; Sitch et al., 2007; Cox et al., 2008; Mercado et al., 2009).

As well as the initialisation of the prognostic variables within the JULES model (Table 2), ancillary information is required for various soil parameters (Table 3). These data are required for both stand alone and coupled applications. In addition, information on the various parameters used within the JULES model is contained in the user documentation, which is attached as supplementary material to this paper.

JULES has been designed to be a flexible modelling system with a modular structure. This structure is illustrated in Fig. 1, where the connections between the modules show the physical processes that connect these areas. The aim of this modular structure is to make it easy to replace modules or to introduce new modules within the modelling system. For instance, whilst at present JULES can be coupled to an external river flow model via the surface and sub-surface runoff fluxes to simulate river discharge, future versions of JULES will include these processes as new modules, along with other processes such as irrigation and groundwater.

Within the modules there are also various science options (Table 4), which can be selected through a series of switches. In general the options represent subsequent developments and improvements to the physics represented in the model. The use of the scheme within an operational weather forecast model (and its evolution from the MOSES land model which was also used in the same environment) requires that such developments are not just simply replaced, but made available as options to ensure backwards compatibility between model versions. However, this presents an opportunity to analyse how developments have impacted the subsequent performance of such a land model.

In addition to the main science modules within JULES there are also three themes. These themes are not connected by physical processes to the other modules, but do impact on each of them and are critical to ensure that the JULES modelling system remains a flexible, easy to use and develop, openly validated tool that can have identifiable configurations for applied applications. These themes include the technical design of the modelling system, the validation and calibration of all aspects of the model, and setting configurations of the modelling system that are suitable for climate impact studies. The themes surround the science modules in Fig. 1 demonstrating their integrating nature.

This paper, the first of two parts that describe the JULES system, is concerned with the energy and water cycles. The second part describes the additional modules required to represent the carbon cycle (Clark et al., 2011), whilst a companion paper addresses one of the cross cutting themes with benchmarking (Blyth et al., 2011). The sections of this paper describe the modules in Fig. 1 relating to energy and water. Section 2 describes the surface exchange, covering (2.1)

Table 2. Prognostic variables within the JULES model.

Data	Units
Fractions of land surface types within gridbox	
Surface temperature of land surface types	K
Temperature of each soil level	K
Moisture concentration of each soil layer	$m^3\,m^{-3}$
Canopy water for vegetation surface types	$kg\,m^{-2}$
Canopy height for vegetation surface types	m
Leaf area index for vegetation surface types	$m^2\,m^{-2}$
Soil carbon	$kgC\,m^{-2}$
Snow amount of each surface type	$kg\,m^{-2}$
Snow grain size on each surface type	µm
Snow on ground below vegetation surface types	$kg\,m^{-2}$
Snowdepth of each surface type[1]	m
Temperature of each snow layer for each surface type[1]	K
Ice content of each snow layer for each surface type[1]	$kg\,m^{-2}$
Liquid content of each snow layer for each surface type[1]	$kg\,m^{-2}$
Snow grain size of each snow layer for each surface type[1]	µm
Mean water table depth[2]	m
Soil moisture in deep layer as a fraction of saturation[2]	

[1] Only for the multi-layer snow option.
[2] Only for the TOPMODEL soil moisture heterogeneity option.

the energy balance equations, (2.2) the surface resistance of moisture for vegetation, (2.3) evaporation of moisture on the surface in either liquid or solid states, (2.4) how urban areas are represented, and (2.5) the treatment of surface heterogeneity.

Section 3 decribes the processes relating to snow. This includes (3.1) the interaction of snow with vegetation canopies, two methods for modelling the snow on the ground, with either (3.2) zero layer or (3.3) multi-layer models, and (3.4) the representation of snow albedo.

Section 4 deals with soil processes for temperature and moisture. This includes (4.1) the amount of water that reaches the soil surface through vegetation canopies and how this is then distributed into runoff and infiltration, (4.2) how soil moisture is extracted from the soil profile by vegetation, (4.3) the thermodynamics and water transport within the soil, (4.4) the hydraulic and (4.5) thermal characteristics of the soil, (4.6) the treatment for preventing a soil layer from becoming super-saturated, and finally (4.7) the representation of heterogeneity for soil moisture. This is done via two possible methods, the first (4.7.1) being based upon the TOPMODEL approach (Beven and Kirkby, 1979), and the second (4.7.2) the PDM model (Moore, 1985).

2 Surface fluxes and energy balance

The surface fluxes of heat, moisture and momentum are calculated in JULES within the surface exchange module. To give the maximum flexibility in terms of the representation of surface heterogeneity and for the coupling of the land surface scheme to an atmospheric model, two generic types of surface are considered; vegetated and non-vegetated. The main difference between these two types of surface is the way in which the surface related parameters (e.g., albedo, roughness length) are specified. For non-vegetative surfaces they are specified by the user (with the exception of the MORUSES option for an urban surface, see Sect. 2.4), whereas for vegetated surfaces these parameters are derived from the structure of the vegetation itself. This leads to an alternative set of parameters that needs to be specified (e.g., rate of change of surface albedo with leaf area index, rate of change of roughness length with canopy height).

2.1 Surface exchange equations

The standard surface energy balance equations, used to calculate the distribution of available energy between the various fluxes at the surface, have been extended to provide more flexibility to include additional physical processes. Thermal inertia is associated with the surface mass which is coupled to the underlying soil by three physical mechanisms depending upon the type of surface. The vegetation fraction is coupled to the soil using radiative exchange and atmospheric turbulence, whereas the remainder are coupled through conduction. The surface energy balance equation is then written:

$$C_s \frac{\delta T_*}{\delta t} = (1-\alpha) S w_\downarrow + \epsilon L w_\downarrow - \sigma \epsilon (T_*)^4 - H - L_c E - G \quad (1)$$

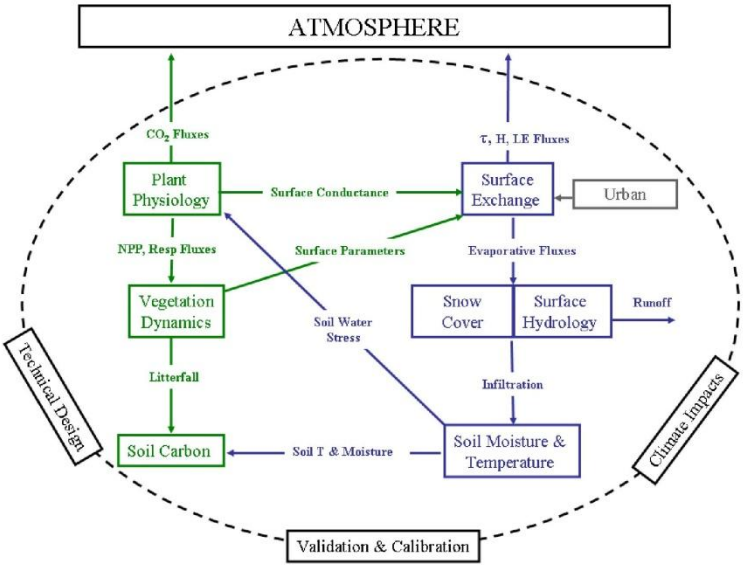

Fig. 1. Modular structure of the JULES model. The boxes show each of the physics modules whilst the lines between the boxes show the physical processes that connect these modules. The surrounding three boxes show the cross-cutting themes.

where:

$$H = \frac{\rho c_p}{r_a}(T_* - T_A) \quad (2)$$

$$E = \frac{\rho}{r_a + r_s}(Q_{sat}(T_*) - Q_1) \quad (3)$$

$$G = \nu\left[\sigma\epsilon\epsilon_s(T_*)^4 - \sigma\epsilon\epsilon_s(T_{s1})^4 + \frac{\rho c_p}{r_{a_{can}}}(T_* - T_{s1})\right] + (1-\nu)\lambda_{soil}(T_* - T_{s1}) \quad (4)$$

The definitions for all symbols are given in Appendix A, along with their units.

For the longwave radiative exchange between vegetation and the soil, one reflection of the emitted radiation is modelled (hence the reason why both emissivities appear in Eq. 4). This assumes that further reflections can be neglected.

A number of options can be chosen to adjust the formulation of the surface energy balance equations. These options increase the level of complexity for the interaction between the surface and the underlying soil, but have the capability to give improvements to the representation of the surface exchange of fluxes and the surface temperature, especially at night (Best and Hopwood, 2001). The traditional surface energy balance equations can be obtained by setting the surface heat capacity to zero (i.e., setting the left-hand side of Eq. (1) to zero) and having only conductive coupling between the surface and the underlying soil (i.e., by setting the vegetation fraction variable to zero in Eq. 4). This was the original surface energy balance that was used within the MOSES model, but Best and Hopwood (2001) showed that this did not provide sufficient cooling during the night over a grass surface. A better fit to the data was given if the surface is radiativly coupled to the underlying soil rather than coupling through conduction. These improvements are provided by the second option which uses not only radiative coupling, but also turbulence between the canopy and the underlying soil for vegetation surfaces, but still retains a zero surface heat capacity ($C_s = 0$). A third option utilises the full energy balance equations above (Eqs. 1–4). This introduces a heat capacity for the surface, which not only gives further improvements for tall vegetation such as forests that have a larger heat capacity than the grass surface considered in (Best and Hopwood, 2001), but also enables other surfaces (such as urban, see section 2.4) to be easily introduced within the model framework. The surface heat capacity is specified for non-vegetation surfaces, but is determined from the leaf and woody biomass for vegetation using

$$C_s = C_L B_L + C_W B_W \quad (5)$$

Larger heat capacities result in a stronger thermal inertia for the surface.

Table 3. Soil ancillary data required by the JULES model.

Data	Units
Bare soil albedo	
Dry soil thermal conductivity	$\mathrm{W\,m^{-1}\,K^{-1}}$
Dry soil thermal capacity	$\mathrm{J\,K^{-1}\,m^{-3}}$
Volumetric saturation point for soil	$\mathrm{m^3\,m^{-3}}$ of soil
Critical volumetric soil moisture content	$\mathrm{m^3\,m^{-3}}$ of soil
Volumetric wilting point for soil	$\mathrm{m^3\,m^{-3}}$ of soil
Soil Saturated hydraulic conductivity of soil	$\mathrm{kg\,m^{-2}\,s^{-1}}$
Saturated soil water pressure (used only for the Brooks and Corey, 1964, soil hydaulics)	m
Clapp-Hornberger exponent (used only for the Brooks and Corey, 1964, soil hydaulics)	
$1/\alpha$ (used only for the van Genuchten, 1980, soil hydaulics)	
$1/(n-1)$ (used only for the van Genuchten, 1980, soil hydaulics)	

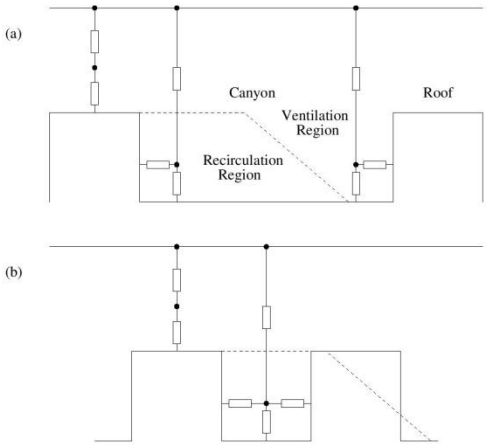

Fig. 2. The two dimensional canyon geometry used in MORUSES illustrating the resistance network used in the parametrisation of the roughness length for heat. The panels depict: **(a)** a wide canyon geometry with both ventilation and recirculation regions; and **(b)** a narrow canyon with only a recirculation region (adapted from Fig. 3 of Harman et al., 2004)

In addition to utilising the full energy balance equations, there is a fourth option which adjusts how snow is represented on vegetation by enabling the snow to exist below the canopy (see Sect. 3.1).

In order to obtain a fully implicit solution, each of the prognostic terms in the surface flux equations (apart from the soil temperature) are written in the form $X^{i+1} = X^i + \Delta X$. The equations are then linearised by assuming that $\Delta X \ll X$. This gives a new set of surface flux equations that can be written in the form of a fully explicit flux, an update to give an implicit solution and a further update to ensure that the atmospheric temperature and humidity satisfy implicit coupling with the atmosphere. The last update is only applied if JULES is connected to an atmospheric model with implicit coupling. So, for example, the surface moisture flux equation becomes:

$$E = \frac{\rho}{r_a + r_s}\left(Q_{sat}\left(T_*^i\right) - Q_1^i\right) + \frac{\rho}{r_a + r_s}\alpha' \Delta T_* - \frac{\rho}{r_a + r_s}\Delta Q_1 \quad (6)$$

where α' is evaluated at T_*^i. The implicit update to the fluxes comes from solving the surface flux equations, whilst the implicit coupling to the atmosphere comes from the coupling methodology of Best et al. (2004).

The aerodynamic resistance is calculated using standard Monin-Obukhov similarity theory (Monin and Obukhov, 1954), using the stability functions of Dyer (1974) for unstable conditions and Beljaars and Holtslag (1991) for stable conditions. The surface resistance for surfaces with potential evaporation (i.e., lake, snow and ice surfaces) is set to zero, whilst for an urban surface the conductance is set to zero unless water is available on the urban surface (i.e., the urban "canopy water"). For a bare soil surface, the surface conductance (g_{soil}, inverse of resistance) is determined by the soil moisture concentration in the top soil layer:

$$g_{soil} = \frac{1}{100}\left(\frac{\theta_1}{\theta_c}\right)^2 \quad (7)$$

This parametrisation was developed following problems identified with a previous scheme (Taylor and Clark, 2001). A review of bare soil evaporation (including Mahfouf and Noilhan, 1991) along with observational studies was used to develop this relationship whilst maintaining consistency with the critical point defined for vegetation (see Sect. 2.2, C. Taylor, personal communication, 2011).

For vegetation, the surface resistance is calculated using the photosynthesis model described in Sect. 2.2.

For the vegetative surfaces, the latent heat flux is determined from a combination of evapotranspiration and bare soil evaporation. The relative contributions from vegetation and bare soil are a representation of the fraction of bare soil that can be seen through the vegetation canopy. Hence the fractions for each of these is determined by the density of the leaves, through the leaf area index. The combined flux represents the interaction of the atmosphere with both the canopy and the soil beneath.

Note this is different to the approach used to represent the evaporation from a sparse canopy. In this situation, due to the limitations of the tile scheme approach as used in JULES (see Sect. 2.5), the surface is distributed into a vegetation land fraction that contributes to a vegetation tile, and a bare soil land fraction that contributes to the bare soil tile.

2.2 Photosynthesis and stomatal conductance

The leaf level stomatal conductance (g_s) and net photosynthetic uptake (A) are linked via the CO_2 diffusion equation:

$$A = g_s(C_c - C_i)/1.6 \tag{8}$$

A second equation by Jacobs (1994), which shares similarities with the simplified form of the Leuning (1995) stomatal conductance formulation, relates the ratio of internal to external CO_2 concentrations to leaf humidity deficit,

$$\frac{C_i - C_*}{C_c - C_*} = f_o\left(1 - \frac{D}{D_*}\right) \tag{9}$$

where f_o and D_* are vegetation specific calibration parameters, which are directly related to the parameters from the Leuning (1995) model (for details, see Cox et al., 1998). This simplified formulation is convenient for large scale model applications (Cox et al., 1998). Potential (non-water stressed) leaf level photosynthesis (A_P) is calculated in JULES using the C3 and C4 photosynthesis models of Collatz et al. (1991) and Collatz et al. (1992) respectively. Photosynthesis is simulated as the minimum of three limiting rates: (i) Rubisco limited rate (W_C), (ii) light limited rate (W_L) and (iii) rate of transport of photosynthetic products (in the case of C3 plants) and PEP-Carboxylase limitation (in the case of C4 plants) W_E. With both, W_C and W_L having a dependency on the leaf internal CO_2 concentration, C_i.

$$A_P = \min(W_C, W_L, W_E) \tag{10}$$

Leaf photosynthesis A, is related to the potential (non-stressed) leaf photosynthesis (A_P) as follows,

$$A = A_P\beta \tag{11}$$

β is the dimensionless moisture stress factor, which is related to the mean soil moisture concentration in the root zone, and the critical and wilting point concentrations as follows:

$$\beta = \begin{cases} 1 & \text{for } \theta \geq \theta_c \\ \frac{\theta - \theta_w}{\theta_c - \theta_w} & \text{for } \theta_w < \theta < \theta_c \\ 0 & \text{for } \theta \leq \theta_w \end{cases} \tag{12}$$

The critical point is defined by a matrix water potential of -33 kPa (Cox et al., 1999), which compares to the more commonly used field capacity that has a matrix water potential of -10 kPa. The use of the critical point enables vegetation to maintain an un-water stressed transpiration at values below field capacity.

JULES uses either a big leaf or a multi-layer approach to scale photosynthesis and conductance to the canopy level. In the big leaf approach, canopy level photosynthesis and conductance are calculated using leaf level fluxes and total canopy leaf area index (Cox et al., 1998) using Beer's law (Monsi and Saeki, 1953). This is the original method used in JULES, but does not produce a realistic dirunal cylce of photosynthesis and hence evaporation (Mercado et al., 2007, 2009). A more realistic scheme is provided by the multi-layer approach, in which the radiation absorbed and photosynthesis are estimated using a user defined number of leaf area increments (canopy layers) within the canopy, with the total canopy level flux calculated as the sum of the fluxes from each individual canopy layer (Jogireddy et al., 2006; Mercado et al., 2007). A number of options are available in JULES for use with this multilayer approach. In addition to the user specifying the number of layers, a two layer approximation can also be selected. This option is not as accurate as the full multilayer scheme, but saves on computational time which can be important for weather forecasting applications. Another option also allows for the variation of leaf nitrogen within the vegetation canopy, leading to further improvements within the multilayer scheme. Equations describing the biochemistry of leaf level photosynthesis (W_C, W_L and W_E) and scaling up methods from leaf to canopy level are outlined in Part II, which describes the carbon cycle in JULES (Clark et al., 2011).

2.3 Freely evaporating surfaces

Evaporation from the surfaces represented within JULES comes from a number of sources. These include evapotranspiration (i.e., water extracted from the soil through vegetation) and bare soil evaporation, both of which include a surface resistance that represents the restrictions in availability of water at the surface. The other sources of evaporation come directly from moisture stores and hence have no surface resistance. These sources include evaporation from open water surfaces, evaporation from surface water held in the canopy of vegetation or ponding on urban surfaces, and sublimation from snow.

The evaporation from water held on the leaves within the vegetation canopy will deplete the canopy water store and can result in all of the water being removed within a timestep. If this occurs, then the moisture unlimited evaporation is set to the available canopy water, and any additional evaporation then comes through evapotranspiration with an associated stomatal, or surface, resistance. Such a limitation in the evaporative flux changes the surface energy balance equations, so an adjustment is made to each of the terms in the energy balance equations to ensure that the model has a consistent solution.

Each surface type within JULES can have snow on it. When snow is present, the surface resistance is set to zero to represent the fact that there is a moisture source. Within JULES there is also an option to have the snow lying underneath vegetation for the turbulent moisture flux (Sect. 3.1). In this case, an additional aerodynamic resistance is added to represent the efficiency of the turbulence at transporting moisture through the canopy. Any sublimation that occurs from the snow on the surface is used to deplete the snow mass in an analogous way to the canopy water. Also like the canopy water, if the snow is removed within a timestep, then an adjustment is made to the terms in the surface energy balance equations to ensure consistency.

Within JULES, lakes can be represented in two ways through the choice of available parameters. The default setting represents lakes as a bare soil surface, except that the surface resistance for the turbulent moisture flux is set to zero, giving a freely evaporating surface. The second method makes use of the surface canopy in the energy balance equations by setting a suitably large value for the surface heat capacity (typically equivalent to water of a depth of around 1 metre, although this can be altered by the user). This option reduces the diurnal cycle of the lake surface temperature compared to the first option, giving a more realistic simulation.

For both methods, as the lake is not explicitly modelled, the evaporative flux is not removed from any moisture store within the model, since it is assumed that there is sufficient water within the lakes to ensure that they are maintained. Similarly, any precipitation that falls onto the lake surface does not contribute to any water store. This means that in order to maintain a water balance, the integrated evaporative flux from the lake surface must be determined and included in the balance equations. This is not routinely done within JULES and has to be calculated through the available diagnostics by the user.

Similarly, the permanent ice surface does not have a prognostic water store, and hence care is required to maintain water balance. To represent an ice surface in JULES, the soil temperature profile is adopted to represent the thermal structure of the ice, whilst the moisture transport used in the soil scheme is neglected. As the ice surface is taken to be one of the surface tile types, and all surface tiles share the same soil information for temperature and moisture in JULES (see Sect. 2.5), this means that it is not possible to have a fractional coverage of land ice within a gridbox or source area at present. As such, there has to be either 100 % of land ice cover or none. The specification of this fraction of land ice is therefore done through the tile fractions information.

As with snow cover (Sect. 3), the surface temperature of the ice surface is prevented from rising above the melting point of water, with any resulting residual of the surface energy balance being added to the melt flux. This means that care must be taken when setting land ice within the JULES model, especially when coupled to an atmospheric model. Small areas of ice could result in large horizontal thermal gradients in the atmosphere, caused by this restriction on the surface temperature compared to ice-free land. This can result in unrealistic small scale circulations and ultimately numerical problems. Hence when coupled to an atmospheric model, this surface type should only be used to represent a large extent of permanent land ice.

2.4 Representation of urban areas

The nature and design of urban environments make their surface energy balance significantly different from natural surfaces. However, a simple bulk representation for an urban area can be obtained by introducing a suitably large thermal capacity for the surface, along with radiative coupling between the surface and the underlying soil. Best (2005) showed that such a simple representation can lead to significant improvements within numerical weather prediction models. The advantage of this approach is that it is easy to adopt within a tile scheme approach and can fit within JULES by adapting currently available parameters.

A second option to represent urban areas in JULES, is to use an additional surface tile. Best et al. (2006) showed that representing the roofs of buildings as one surface and street canyons as a second effective surface gives improvements over the one-tile approach. Also, Harman and Belcher (2006) and Porson et al. (2009) demonstrated that these two surfaces give a good approximation of more complex schemes that represent each of the facets within the urban area. The differences between the two surface types is given through the surface parameter specifications.

The third option implemented is the Met Office Reading Urban Surface Exchange Scheme (MORUSES), as described in Porson et al. (2010a,b). Again this is a two-tile scheme, but as the surface parameters are determined from the morphology and material properties of the city, this enables a distribution of surface fluxes with different structural properties. The radiative exchange within the canyon tile is formulated with an effective albedo and an effective emissivity, based upon the exchanges between the various street canyon facets. The roughness length for momentum for the urban area is determined from the formulation of Macdonald et al. (1998), for a staggered array of cubes; the canyon and the roof tiles both have the same roughness.

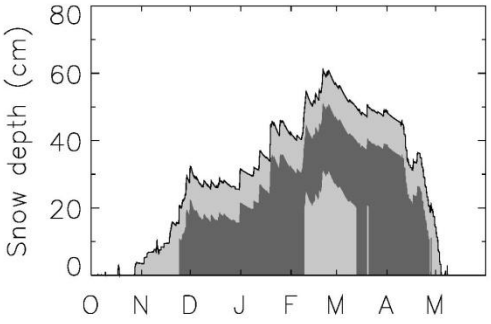

Fig. 3. Simulation of seasonal snow depth with JULES for multi-level snow scheme, showing the division into a varying number of layer depths. The full shaded area shows the total snow depth, whilst the different shadings represent the depths of the various snow layers. Minimum layer thicknesses can be selected by the user, but in this illustration a second layer is added when the snow depth exceeds 20 cm and a third at 30 cm.

The roughness length for temperature comes from a physically-based parametrisation that relates to the urban morphology and uses a resistance network to represent the transfer of heat (see Fig. 2). The canyon tile includes the effects of the recirculation jets by using two resistance pathways; one for each of the recirculation and ventilation regions. For both of these elements, three resistances are used, two representing the heat across an internal boundary layer adjacent to each facet and one representing the transfer of heat across the inertial sub-layer. The roof, which is simpler, only has two resistances representing the internal boundary layer and inertial sub-layer (see Harman et al., 2004; Porson et al., 2010a, for more details).

Effective areal heat capacities are determined to represent the roof and the canyon, which includes contributions from both the walls and the road. These are determined by considering the diurnal response using a force-restore model, whilst an adjustable roof parameter is also introduced to increase the flexibility to capture different oscillations. The canyon tile is conductively coupled through the road to the underlying soil surface, whilst the walls of the canyon and the roof tile are decoupled from the soil, by imposing a zero flux boundary condition.

The data MORUSES requires can be sourced from a variety of places, depending on availability. For example MORUSES has been used to simulate the London urban heat island as part of the LUCID project (The Development of a Local Urban Climate Model and its Application to the Intelligent Design of Cities; Bohnenstengel et al., 2011) in which the Virtual London model (Evans et al., 2005) was used for building geometry. As part of the LUCID project, empirical relationships were also formulated for urban geometry to represent areas within the study area that did not exist within the Virtual London domain. The building material properties used were typical values: clay roof, brick walls and asphalt road. Where no information of this kind is known by the user, a global dataset also exists that categorises urban areas depending on density, climatic conditions and regional culture (Jackson et al., 2010). However, the amount and quality of the data known by the user would ultimately govern the choice of urban model used.

2.5 Surface heterogeneity

The heterogeneity of the surface is modelled within JULES by using the tile, or mosaic, approach (e.g., Essery et al., 2003a). This means that a separate surface energy balance is determined for each type of surface within the domain of the gridbox or footprint, and the individual surface fluxes are then given a weighted average in order to determine the gridbox or footprint mean flux into the atmosphere. One limitation to the current structure of JULES is that although the surface exchange represents the heterogeneity through tiling, there is no representation of sub-gridscale heterogeneity within the sub-surface soil module. This will be developed in future versions of the JULES model.

In order to keep the parametrisation of surface heterogeneity as flexible as possible, the number of surface types to be considered within a model simulation is determined at run time. Hence the complexity of the heterogeneity and cost in terms of computational time have to be balanced. Thus a time-limited modelling application, such as operational weather forecasting, can run with minimal surface types to optimize cost, whereas other applications may benefit from unlimited surface types (e.g., climate applications with an interactive carbon cycle).

There are two generic types of surface in JULES having differing requirements for their surface parameters: (1) Non-vegetated surfaces with fixed parameter values (e.g., albedo and roughness length) which are specified at run time, and (2) vegetated surfaces whose parameters vary. The latter are described in the following paragraphs.

The roughness length for momentum for vegetation is determined from

$$z_0 = \omega h \tag{13}$$

There are two options to determine the surface albedo (α) for vegetation. The simplest option is a bulk albedo:

$$\alpha = \alpha_b \exp(-kL) + \alpha_\infty \left[1 - \exp(-kL)\right] \tag{14}$$

where α_b, the soil albedo, is a spatially varying ancillary field within JULES and α_∞ is the prescribed maximum canopy albedo for dense leaf coverage.

With the second option, the snow-free albedos are calculated using the two-stream model for radiative transfer through vegetation described by Sellers (1985). This scheme

uses separate direct-beam and diffuse albedos in the visible and near-infrared wave bands for each vegetation type. This requires four parameter values for leaf reflection coefficients and leaf scattering coefficients for both near infra-red and photosynthetically active radiation.

An additional parameter for vegetation surfaces is the capacity of the canopy to hold water (C_m) through the interception of precipitation,

$$C_m = A_m + B_m L \tag{15}$$

By default nine surface types are represented; five vegetation (broadleaf trees, needleleaf trees, C3 grasses, C4 grasses and shrubs) and four non-vegetated surfaces (urban, open water, bare soil and permanent land ice). The default parameters for each of these surface types are given in Tables 5 and 6, but where possible these parameters should be calibrated for specific sites.

In addition to the surface type, each tile has an elevation above the mean gridbox height. This enables surfaces that are sensitive to the changes in atmospheric temperature and humidity, arising from displacement above the mean surface height, to experience adjusted atmospheric forcing. This is done in a simple way by adjusting the air temperature along a dry adiabat whilst keeping the specific humidity constant until the saturation point is reached. After this, the temperature is adjusted along a moist adiabat, whilst the specific humidity is then set to the saturated specific humidity at the new atmospheric temperature. To ensure consistency with the top soil level temperature, this is adjusted by the same increment as the air temperature during the calculation of the surface energy balance (note that the actual prognostic soil temperature variable is not updated by this increment). This prevents artificial warming from the soil without having to introduce heterogeneity into the soil. This assumption will be removed once soil heterogeneity is introduced into the JULES code. One impact of introducing elevation bands is to reduce spurious sublimation and melting from snow-covered surfaces.

3 Snow model

Two schemes are available within JULES for the representation of snow on the ground. The simplest is a zero-layer scheme that uses no explicit model layers to represent snow, instead adapting the top soil level to represent lying snow processes. The more comprehensive and physically realistic scheme takes a multi-layer approach. For vegetated surfaces, snow may additionally be partitioned between intercepted snow in the canopy and snow on the ground or held in a single effective store. The simple and multi-layer snow scheme give similar results in many conditions, but the multi-layer scheme is expected to give better simulations of snow dynamics at sites with deep snow, with the possibility of mid-winter melt events and better simulations of soil temperatures at sites with low winter air temperatures.

3.1 Interaction of snow with vegetation canopies

With the original scheme in MOSES, snow is held in a single store and hence sits on top of vegetation regardless of the type and height of this vegetation. The exception of this is for the albedo which does account for a darker surface when snow is under tall vegetation. However, PILPS-2e (Bowling et al., 2003) found that the models with highest winter sublimation had lowest annual runoff for a high-latitude basin. A reformulation of MOSES to distinguish between snow on and below forest canopies reduced the sublimation and improved the simulation of runoff (Essery and Clark, 2003). So an option is also available in JULES to partition the snow between snow on the canopy and the underlying ground (Essery et al., 2003a). The surface resistance for sublimation is set to zero for tiles with snow cover in the single-store option, but is

$$r_s = \frac{\rho_i r^2}{0.03\, D_f (1.79 + 3\, U^{1/2})\, I} \left(\frac{I}{I_{max}} \right)^{0.4} \tag{16}$$

for canopy snow, where $I_{max} = 4.4L$ is the snow interception capacity for a canopy with leaf area index L and $r = 0.5$ mm is a nominal grain radius for intercepted snow (Essery et al., 2003b). The change in load during a timestep with snowfall amount S_f on a canopy with initial load I_0 is

$$\Delta I = 0.7\,(I_{max} - I_0)\left(1 - e^{-S_f/I_{max}}\right). \tag{17}$$

Snow is removed from the canopy by sublimation, and unloading of melting snow from the canopy is set equal to 40 % of the canopy snowmelt rate (Storck et al., 2002; Essery et al., 2003b).

3.2 Zero-layer snow model

The original snow scheme within JULES is a zero-layer snow model. Snow is given a constant thermal conductivity and a constant density. The heat capacity of snow is neglected, but snow decreases the bulk thermal conductivity of the surface layer due to both the increased layer thickness and the different conductivities of snow and soil. For snow depth less than half the surface soil layer thickness (Δz_1), the thermal conductivity used in surface energy balance calculations is adjusted for insulation by snow according to

$$\lambda = \lambda_{soil} \left[1 + \frac{2 d_s}{\Delta z_1} \left(\frac{\lambda_{soil}}{\lambda_{snow}} - 1 \right) \right]^{-1} \tag{18}$$

The heat flux between the surface layer and the second soil layer, of thickness Δz_2, is multiplied by a snow insulation factor

$$\zeta = \left(1 + \frac{2 d_s}{\Delta z_1 + \Delta z_2} \right)^{-1}. \tag{19}$$

Table 4. Description of the various physics options within the JULES model as discussed in the identified sections.

Physics	Section	Option
Surface Exchange	2.1	No thermal inertia and conductive coupling
		No thermal inertia and radiative coupling
		Thermal inertia and radiative coupling
		Thermal inertia, radiative coupling and snow under vegetation canopy
Canopy radiation and scaling	2.2	Big leaf
		Multiple canopy layers
		2 layer approximation to multiple canopy layers
		Multiple canopy layers with variable leaf nitrogen
Albedo	2.5	Bulk albedos
		Spectral albedos and snow ageing
Urban model	2.4	1 tile (bulk)
		2 tiles (roofs and canyons)
		MORUSES; Porson et al. (2010a,b)
Snow	3.2	zero layer model
	3.3	multi-layer model
Soil hydraulics	4.4	Brooks and Corey (1964)
		van Genuchten (1980)
Soil thermodynamics	4.5	Cox et al. (1999)
		Dharssi et al. (2009)
Soil moisture super-saturation	4.6	Restricted drainage into layer
		Infiltration into lower layer
Large scale hydrology	4.7.1	TOPMODEL Gedney and Cox (2003)
	4.7.2	PDM (Moore, 1985)

For deeper snow, the surface conductivity is set equal to $\lambda_{\rm snow}$ and the insulation factor is

$$\zeta = (\Delta z_1 + \Delta z_2)\left[(2d_{\rm s} - \Delta z_1)\frac{\lambda_{\rm soil}}{\lambda_{\rm snow}} + 2\,\Delta z_1 + \Delta z_2\right]^{-1} \quad (20)$$

(Cox et al., 1999). The surface skin temperature is not allowed to exceed 0°C while snow remains on the ground, and the heat flux used to melt snow is diagnosed as a residual in the surface energy balance. Melt water drains immediately from the snow and is partitioned into soil infiltration and runoff; there is no storage or freezing of liquid water in snow. The snow thermal conductivity, snow density and surface layer thickness are parameters that are set by the user.

Whilst the zero-layer snow scheme on the whole gives good agreement with observations, it tends to melt snow too rapidly. This is partly due to the inability of this scheme to hold liquid water within the snow that can subsequently re-freeze. In addition, the use of the top soil layer to represent the snow has a negative impact on the soil temperatures, e.g., as demonstrated in the SNOWMIP2 experiment (Essery et al., 2009). Better agreement between observations and the JULES model can be obtained by using the alternative multi-layer snow scheme.

3.3 Multi-layer snow model

The maximum number of layers ($N_{\rm max}$) that are used for deep snow and their thickness d_k (k=1,...,$N_{\rm max}$) are set by the user. However, the number of layers actually used depends on the snow depth, which means that not all the layers exist at any one time. When a layer is at the base of the snowpack it has a variable thickness. Shallow snow is combined with the surface soil layer for snow depth $d_{\rm s} < d_1$ for numerical stability, whilst setting $N_{\rm max} = 0$ forces the use of the zero-layer option for any depth of snow. For $d_{\rm s} \geq d_1$, snow is represented by additional model layers on top of the soil if $N_{\rm max} \geq 1$. As the snow depth increases, the lowest layer in the snowpack increases in thickness until it reaches twice its prescribed thickness; the layer then splits in two with the upper part staying fixed in thickness and the new lowest layer thickening as the snow accumulates. This is reversed as the snow depth decreases, with layers being progressively combined at the bottom of the snowpack. The division of a snowpack into layers is illustrated in Fig. 3. A variable snow density is used, so snow depth can decrease due to compaction as well as ablation.

Each layer in the snowpack has a thickness d_k (m), a temperature T_k (K), a density ρ_k (kg m^{-3}), an ice content I_k (kg m^{-2}) and a liquid water content W_k (kg m^{-2}). Layer thickness, density and mass are related by $\rho_k d_k = I_k + W_k$. The increase in layer density due to compaction over a timestep of length δt is calculated as

$$\frac{\delta \rho_k}{\delta t} = \frac{\rho_k\, g\, M_k}{\eta_0} \exp\left(\frac{k_{\rm s}}{T_{\rm m}} - \frac{k_{\rm s}}{T_k} - \frac{\rho_k}{\rho_0}\right) \quad (21)$$

Table 5. Default parameter values required by JULES for the standard vegetation surfaces.

Parameter	Broadleaf trees	Needleleaf trees	C3 grasses	C4 grasses	Shrubs
Snow-covered albedo for large LAI	0.15	0.15	0.60	0.60	0.40
Snow-covered albedo for zero LAI	0.30	0.30	0.80	0.80	0.80
Snow-free albedo for large LAI	0.10	0.10	0.20	0.20	0.20
Rate of change of vegetation roughness length with height	0.05	0.05	0.10	0.10	0.10
Minimum canopy capacity ($kg\,m^{-2}$)	0.50	0.50	0.50	0.50	0.50
Rate of change of canopy capacity with LAI	0.05	0.05	0.05	0.05	0.05
Infiltration enhancement factor	4.00	4.00	2.00	2.00	2.00
Light extinction coefficient	0.50	0.50	0.50	0.50	0.50
Rootdepth (m)	3.00	1.00	0.50	0.50	0.50

where $k_s = 4000\,K$, compactive viscosity $\eta_0 = 10^7\,Pa\,s$, reference density $\rho_0 = 50\,kg\,m^{-3}$, temperature $T_m = 273.15\,K$, and $M_k = 0.5(I_k + W_k) + \sum_{i=1}^{k-1}(I_i + W_i)$ is the mass of snow above the middle of the layer. This scheme, based on measurements by Kojima (1967), has previously been used in the snow models described by Pitman et al. (1991) and Lynch-Stieglitz (1994). The areal heat capacity of a layer is

$$C_k = I_k\,C_{ice} + W_k\,C_{water} \tag{22}$$

where $C_{ice} = 2100\,J\,K^{-1}\,kg^{-1}$ and $C_{water} = 4180\,J\,K^{-1}\,kg^{-1}$ are the specific heat capacities of ice and water, and the thermal conductivity is

$$\lambda_k = 2.22\left(\frac{\rho_k}{\rho_{water}}\right)^{1.88} \tag{23}$$

where $\rho_{water} = 1000\,kg\,m^{-3}$ is the density of water (Yen, 1981).

The structure of the multi-layer snow model is shown in Fig. 4. The conducted heat flux at the bottom of layer k is

$$H_k = \Gamma_k\,[T_k - T_{k+1} + \gamma\,(\delta T_k - \delta T_{k+1})] \tag{24}$$

where δT_k is the increment in layer temperature over a timestep, γ is the forward timestep weighting (0 for explicit and 1 for fully implicit timestepping), and

$$\Gamma_k = \left(\frac{d_k}{2\lambda_k} + \frac{d_{k+1}}{2\lambda_{k+1}}\right)^{-1} \tag{25}$$

is a layer thickness weighted thermal conductivity. For the lowest snow layer ($k = N$), T_{N+1}, d_{N+1} and λ_{N+1} are the temperature, thickness and conductivity of the surface soil layer. The increment in layer temperature over a timestep is

$$\delta T_k = \frac{\delta t}{C_k}(H_{k-1} - H_k) \tag{26}$$

Surface heat flux H_0 calculated by the surface exchange module is passed to the snow module, and ground heat flux H_N calculated by the snow module is passed to the soil module; implicit timestep weighting of surface soil layer temperature T_{s1} is not used in calculating this flux. For a single snow layer the temperature increment is given by

$$\delta T_1 = \frac{\delta t}{C_1}[H_0 - \Gamma_1(T_1 - T_{s1} + \gamma\,\delta T_1)] \tag{27}$$

with solution

$$\delta T_1 = \frac{[H_0 + \Gamma_1(T_{s1} - T_1)]\,\delta t}{\gamma\,\Gamma_1\,\delta t + C_1} \tag{28}$$

When there are $N > 1$ snow layers, increments in the layer temperatures are found as the solutions of the tridiagonal set of equations

$$b_1\,\delta T_1 + c_1\,\delta T_2 = [H_0 + \Gamma_1(T_2 - T_1)]\,\delta t, \tag{29}$$

$$a_k\,\delta T_{k-1} + b_k\,\delta T_k + c_k\,\delta T_{k+1} = [\Gamma_{k-1}(T_{k-1} - T_k) + \Gamma_k(T_{k+1} - T_k)]\,\delta t \tag{30}$$

for $k = 2, ..., N-1$, and

$$a_N\,\delta T_{N-1} + b_N\,\delta T_N = [\Gamma_{N-1}(T_{N-1} - T_N) + \Gamma_N(T_1 - T_N)]\,\delta t \tag{31}$$

with matrix elements

$$a_k = c_{k-1} = -\gamma\,\Gamma_{k-1}\,\delta t \tag{32}$$

and

$$b_k = C_k + \gamma\,(\Gamma_{k-1} + \Gamma_k)\,\delta t. \tag{33}$$

If the temperature of a layer is calculated to be above T_m, the layer ice mass is reduced by an amount

$$\delta I_k = \frac{C_k}{L_f}(T_k - T_m) \tag{34}$$

or the entire mass of the layer, whichever is least. The layer liquid mass is increased by the same amount and the layer

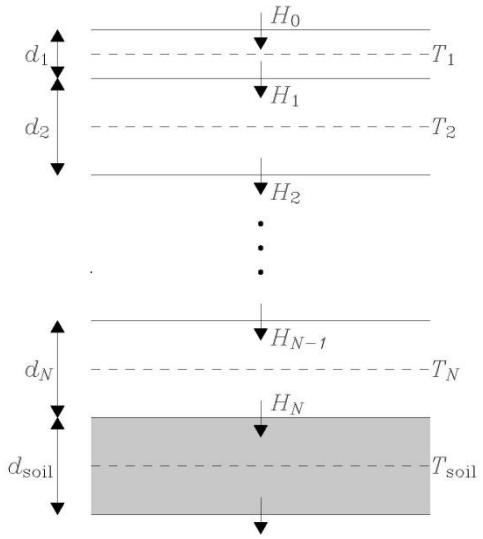

Fig. 4. Structure of the numerical discretisation over the layers for the temperatures and heat fluxes within the multi-level snow scheme in JULES.

temperature is reset to T_m. Sublimation calculated by the surface exchange module is removed from the surface layer ice mass and from deeper layers if the surface layer sublimates entirely during a timestep.

A layer of depth d_k entirely consisting of liquid water would have a liquid content of $\rho_{water}\ d_k$. Snow layers are allowed to retain a fraction W_{cap} (set by the user) of this liquid content. When the liquid content of a layer exceeds its capacity, excess water is passed to the layer below. Liquid water in a layer with temperature below T_m will freeze, decreasing the liquid content by an amount

$$\delta W_k = \frac{C_k}{L_f}(T_m - T_k), \tag{35}$$

increasing the ice content by the same amount, and increasing the temperature by

$$\delta T_k = \frac{L_f \delta W_k}{C_k}. \tag{36}$$

The water flux at the base of the snowpack is passed to the surface hydrology module (Sect. 4.1).

Fresh snow is added as an interim layer 0 with density ρ_0 and temperature equal to the surface layer temperature. After increments have been applied to the layer masses and temperatures, layers are combined or split as necessary to match the fixed layer thicknesses. The liquid water and ice contents of the revised snow layers are determined by conservation of mass and their temperatures are diagnosed from conservation of energy

3.4 Snow albedo

Diagnostic and prognostic snow albedo options are provided. The simpler diagnostic option was originally used in MOSES, but can not represent the impacts of snow ageing on the surface albedo. Hence the prognostic option provides the ability to represent the time evolution of the snow ableldo, improving the physical representation of the snow.

In the diagnostic scheme, a snow-free albedo and an albedo for cold deep snow are specified for each surface type. When the surface temperature exceeds a threshold temperature T_c, the snow albedo is decreased according to

$$\alpha_s = \alpha_{cds} + k_1(\alpha_0 - \alpha_{cds})(T_* - T_c). \tag{37}$$

For a tile with snow depth d_s, the albedo is a weighted average

$$\alpha = \alpha_0 + (\alpha_s - \alpha_0)(1 - e^{-d_s/d_m}) \tag{38}$$

for surface masking snow depth d_m.

For tall vegetation, the impact of snow lying underneath the vegetation canopy is taken into account by setting lower values for the cold deep snow albedo.

The prognostic albedo scheme uses the Wiscombe and Warren (1980) spectral snow model. The ageing of the snow surface is characterized by introducing a prognostic grain size $r(t)$, set to $r_0 = 50\,\mu m$ for fresh snow and limited to a maximum value of 2000 µm. The change in grain size over a timestep is given by

$$r(t+\delta t) = \left[r(t)^2 + \frac{G_r}{\pi}\delta t\right]^{1/2} - [r(t) - r_0]\frac{S_f \delta t}{d_0} \tag{39}$$

for snowfall rate S_f. The mass of fresh snow required to refresh the albedo is set to $2.5\,kg\,m^{-2}$. The empirical grain area growth rate, in $\mu m^2\ s^{-1}$, is

$$G_r = \begin{cases} 0.6 & T_* = T_m \text{ (melting snow)} \\ 0.06 & T_* < T_m,\ r < 150\,\mu m \text{ (cold fresh snow)} \\ A_s \exp(-4550/T_*) & T_* < T_m,\ r > 150\,\mu m \text{ (cold aged snow)} \end{cases} \tag{40}$$

where $A_s = 0.23 \times 10^6\ \mu m^2\ s^{-1}$. Snow albedos for diffuse visible and near-infrared radiation are calculated as

$$\alpha_{vis} = 0.98 - 0.002(r^{1/2} - r_0^{1/2}) \tag{41}$$

and

$$\alpha_{nir} = 0.7 - 0.09 \ln\left(\frac{r}{r_0}\right). \tag{42}$$

The zenith-angle dependence of albedos for direct-beam radiation with zenith cosine μ is represented by using an effective grain size,

$$r_e = [1 + 0.77(\mu - 0.65)]^2 r, \tag{43}$$

in place of r in the equations for diffuse albedos (Eqs. 41 and 42).

For a tile with snow-free albedo α_0, snowdepth d_s and roughness length z_0, the albedo in each band is

$$\alpha = f_s \alpha_s + (1 - f_s)\alpha_0 \tag{44}$$

where

$$f_s = \frac{d_s}{d_s + 10 z_0} \tag{45}$$

(Oleson et al., 2004). When driving data with separate direct-beam and diffuse radiation in visible and near-infrared bands are not available, the average of the diffuse albedos is simply used as an all-band snow albedo.

4 Hydrology and soil thermodynamics

JULES includes multi-layer, finite-difference models of sub-surface heat and water fluxes, as described in Cox et al. (1999). There are options for the specification of the hydraulic and thermal characteristics, the representation of super-saturated soil moisture and the sub-surface heterogeneity of soil moisture.

4.1 Surface hydrology

To account for the size of convective storms compared to gridsize, a rainfall rate is assumed to fall on a fraction ϵ_r of the grid. For large scale precipitation and point studies this fraction is set to one, whilst for convective precipitation it can take lower values, and is typically set to a value of 0.3.

The amount of water that reaches the soil surface depends upon the type of surface. For non-vegetation surfaces, this is simply the precipitation rate whereas for vegetation surfaces, this becomes the throughfall and is calculated using:

$$T_F = R\left(1 - \frac{C}{C_m}\right)\exp\left(-\frac{\epsilon_r C_m}{R\,\Delta t}\right) + R\frac{C}{C_m} \tag{46}$$

and the canopy water is updated by

$$C^{(n+1)} = C^{(n)} + (R - T_F)\,\Delta t. \tag{47}$$

where C_m is the maximum canopy water that can be held by the vegetation and Δt is the timestep.

The canopy water can also be increased through dewfall (i.e., downward surface moisture fluxes), and is depleted by surface evaporation. Similarly, snow cover is increased through the deposition of frost (modelled as dewfall at surface temperatures below freezing), whilst the melting of snow contributes to the water available at the soil surface and updates the equivalent water within the snow pack.

The water reaching the soil surface is then split between infiltration into the soil and surface runoff. Since the throughfall can be different for various surface types within a tile scheme, whereas in JULES the infiltration into the soil is a grid-box aggregate as the soil is not tiled, the surface runoff is determined by combining equations for both the throughfall and the grid-box mean infiltration, yielding the following:

$$Y = \begin{cases} R\frac{C}{C_m}\exp\left(-\frac{\epsilon_t K C_m}{RC}\right) + R\left(1 - \frac{C}{C_m}\right)\exp\left(-\frac{\epsilon_t C_m}{R\,\Delta t}\right) & K\,\Delta t \le C \\ R\exp\left[-\frac{\epsilon_t (K\,\Delta t + C_m - C)}{R\,\Delta t}\right] & K\,\Delta t > C \end{cases} \tag{48}$$

where the surface infiltration rate K is equal to $\beta_s K_{hs}$ and β_s is an enhancement factor. These equations account for the effect of a finite model timestep on the throughfall and therefore the surface runoff. A full derivation of these equations was given by Dolman and Gregory (1992).

The infiltration into the soil is determined through the integration of the contributions for each of the surface types by using the water balance at the surface:

$$W_0 = \sum_j \nu_j\,(T_{Fj} + S_{mj} - Y_j). \tag{49}$$

4.2 Soil moisture extraction

The ability of vegetation to access moisture at each level in the soil is determined by root density, assumed to follow an exponential distribution with depth. The fraction of roots in soil layer k extending from depth z_{k-1} to z_k is

$$r_k = \frac{e^{-2z_{k-1}/d_r} - e^{-2z_k/d_r}}{1 - e^{-2z_t/d_r}}, \tag{50}$$

For transpiration E', the flux extracted from soil layer k is $e_k^0 E'$, where

$$e_k^0 = \frac{r_k\,\beta_k}{\sum_k r_k\,\beta_k} \tag{51}$$

and

$$\beta_k = \begin{cases} 1 & \theta_k \ge \theta_c \\ (\theta_k - \theta_w)/(\theta_c - \theta_w) & \theta_w < \theta_k < \theta_c \\ 0 & \theta_k \le \theta_w \end{cases} \tag{52}$$

is a soil moisture availability factor, defined similarly to Eq. (12), for a soil layer with unfrozen volumetric soil moisture concentration θ_k.

4.3 Soil thermodynamics and water fluxes

The sub-surface at a gridpoint is either soil or land ice (with no water movement in the latter). Sub-surface temperatures are calculated using a finite difference form of the heat diffusion equation, including the effects of solid-liquid phase changes of water. The temperature of the k-th soil layer is incremented by the diffusive heat fluxes into and out of the layer, G_{k-1} and G_k, and the advective flux from the layer by flowing water, J_k:

$$C_a \Delta z_k \frac{dT_{sk}}{dt} = G_{k-1} - G_k - J_k \Delta z_k \quad (53)$$

where the fluxes are calculated as

$$G = \lambda_{soil} \frac{\partial T_s}{\partial z} \quad (54)$$

$$J = C_{water} W' \frac{\partial T_s}{\partial z} \quad (55)$$

where z is the vertical coordinate. C_a is the "apparent" volumetric heat capacity of the layer, including the effect of phase changes (Cox et al., 1999). For soil, the sub-surface thermal characteristics are a function of solid and liquid water contents, while land ice uses fixed characteristics. The top boundary condition for Eq. (53) is the surface heat flux, calculated by the surface exchange module, while at the bottom there is a zero flux boundary condition to ensure conservation of energy.

The number of soil layers is a model parameter but the default is four of thicknesses 0.1, 0.25, 0.65 and 2.0 m, giving a total soil depth of 3 m. This configuration is designed to capture the variation of soil temperature from sub-daily to annual timescales (Best et al., 2005).

Soil water contents are updated using a finite difference form of the Richards equation. The moisture content of each layer is updated as:

$$\frac{d\theta_k}{dt} = W'_{k-1} - W'_k - E'_k - R_{bk} \quad (56)$$

where W'_{k-1} and W'_k are the diffusive fluxes flowing in from the layer above and out to the layer below respectively, E'_k is the evapotranspiration extracted by plant roots in the layer (and bare soil evaporation for the top layer) and R_{bk} is lateral runoff, which is set to zero unless the sub-surface heterogeneity of soil moisture is represented using the TOPMODEL option (Sect. 4.7.1). The vertical fluxes follow Darcy's law

$$W' = K_h \left(\frac{\partial \Psi}{\partial z} + 1 \right) \quad (57)$$

The top boundary condition for Eq. (56) is the infiltration of water at the soil surface, whilst the lower boundary condition is drainage, which contributes to sub-surface runoff.

4.4 Hydraulic characteristics

There are two options for the hydraulic characteristics. In the first the relation between soil water content, suction and hydraulic conductivity are Brooks and Corey (1964):

$$\theta/\theta_s = (\Psi/\Psi_s)^{-1/b} \quad (58)$$

$$K_h = K_{hs}(\theta/\theta_s)^{2b+3} \quad (59)$$

This is the method that has traditionally been used in land surface models. It is a more simplistic formulae for the hydraulic properties of the soil than other schemes, but the required parameters can be determined from the sand, silt and clay fractions of soils, which are available in many global soil datasets.

The parameters θ_s, Ψ_s and b are calculated from soil texture information using the relationships of Cosby et al. (1984) or others. (Note that Cox et al. (1999) incorrectly referenced Eqs. (58) and (59) as Clapp and Hornberger (1978) rather than Brooks and Corey (1964) (T. Marthews, personal communication, 2009).)

The second option uses the hydraulic relationships of van Genuchten (1980), which is a more complex formulae but more scientifically robust:

$$\frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{1}{[1 + (\alpha_v \Psi)^n]^m} \quad (60)$$

$$K_h = K_{hs} S^{\xi} \left[1 - (1 - S^{1/m})^m \right]^2 \quad (61)$$

where $m = 1 - 1/n$ and $S = (\theta - \theta_r)/(\theta_s - \theta_r)$. In JULES, $\xi = 0.5$ and the soil moisture variable is implicitly defined as $\theta - \theta_r$, leaving three parameters. The specific parameters required for this formulation have not been traditionally available within soil dataset, making it difficult to use. However, more recent datasets now include these parameters within their soil information.

Dharssi et al. (2009) show that with suitable parameter values, Eqs. (58) and (60) are similar over most of the soil moisture range.

The soil parameter values can vary between layers but, in the absence of suitable data with which to specify this variation, many applications ignore any variation with depth. When calculating the hydraulic characteristics using Eqs. (58–61), JULES uses θ_u, the unfrozen volumetric water content, instead of θ, to capture the effects of soil freezing, following Cox et al. (1999).

4.5 Thermal characteristics

JULES has two options for calculating the effective thermal conductivity of soil λ. The first option (described by Cox et al., 1999) is a less complex scheme, but requires only a limited amount of soil information:

$$\lambda = (\lambda_s - \lambda_{dry})\theta/\theta_s + \lambda_{dry} \quad (62)$$

where

$$\lambda_s = \lambda_{water}^{\theta_u^s} \lambda_{ice}^{\theta_f^s} \lambda_{dry} / \lambda_{air}^{\theta_s}, \quad (63)$$

where $\theta_f^s = \theta_s[S_f/(S_u + S_f)]$, $\theta_u^s = \theta_s - \theta_f^s$, and S_u and S_f are the unfrozen and frozen water contents as a fraction of saturation.

Dharssi et al. (2009) showed that the thermal characteristics from this scheme do not agree well with those for various soil types. It was found that the formulation of Johansen (1975) gives a better fit, but this scheme requires additional soil information. Although this additional information is

generally available in the latest soil datasets, implementing it into the Met Office Unified Model requires significant effort. Hence Dharssi et al. (2009) derived a simplification of Johansen (1975) which gives similar response in the relationships between the thermal conductivity and soil moisture:

$$\lambda = (\lambda_s - \lambda_{dry}) K_e + \lambda_{dry} \tag{64}$$

where K_e is the Kersten number

$$K_e = \begin{cases} \log(\theta/\theta_s) + 1.0 & (\theta/\theta_s) \geq 0.1 \\ 0 & \text{otherwise} \end{cases} \tag{65}$$

$$\lambda_s = \frac{\lambda_{water}^{\theta_u^s} \lambda_{ice}^{\theta_f^s}}{\lambda_{water}^{\theta_s}} \lambda_s^u \tag{66}$$

$$\lambda_s^u = 1.58 + 12.4(\lambda_{dry} - 0 \cdot 25) \tag{67}$$

with the constraint that $1{\cdot}58 \leq \lambda_s^u \leq 2{\cdot}2$. This equation for the thermal conductivity of unfrozen saturated soil (λ_s^u) was derived in order to give good agreement with the Johansen (1975) formulation, but without requiring knowledge of the mineral content of the soil (Dharssi et al., 2009).

This generally gives larger values for conductivity than the Cox et al. (1999) formulation, which reduces the errors in simulated air temperature when used in Numerical Weather Prediction (Dharssi et al., 2009).

4.6 Super-saturation of soil moisture

The numerical solution for the transport of soil moisture between the soil layers may result in layers which become super-saturated. JULES has two options to prevent this from occurring. With the first option, if a soil layer becomes super-saturated, then the soil moisture in this layer is limited to the saturation point and the excess water is prevented from moving into the layer from above, i.e., the drainage into the layer is restricted by the saturation. This results in the excess water being moved back up the soil layer, and if the top soil layer becomes super-saturated, then the excess water is added to the surface runoff.

The second option is to route the excess soil moisture to the soil layers below. This assumes excess soil moisture might flow laterally over land within a large gridbox, but would eventually move down through the soil layers at sub-grid locations in which drainage is less impeded (e.g., where there is fractured permafrost or less compacted/faster draining soil types). This results in the excess water being moved down to lower layers, and if the bottom soil layer becomes super-saturated, then the excess water is added to the sub-surface runoff.

If the total soil column is saturated, then the difference between these two options is to add the excess water to either the surface or sub-surface runoff. Whilst in both cases the water results in a runoff flux, this could impact the timing of river flow due to the delay of sub-surface runoff getting into the river network. Tests of the two options with the PILPS2d Valdai data (Schlosser et al., 2000) showed that moving the excess water in the downwards direction led to a poor surface runoff simulation and excessive soil moisture, whereas inhibiting the drainage of water into a saturated layer gave a better agreement with observations.

However, global simulations have shown that in regions of partially frozen soils, one possible result is saturated and partially frozen soil layers near the surface, with unsaturated layers below. In this situation, the option to inhibit the drainage of water into the saturated layer at the surface leads to excessive surface runoff of snowmelt, giving a dry soil during spring and hence a dry and warm bias in the atmosphere during the summer. The option to move the excess water to lower layers moistens the lower unsaturated soil layers and removes some of this dry and warm atmospheric bias whilst reducing the surface runoff of snowmelt.

These results suggest that the grid size may be important in determining the dominant physical processes that prevent the super-saturation of the soil, and further work is required to determine how this should be represented in the model. Thus care should be taken when choosing between options for controlling super-saturation, with consideration being taken for the required application.

4.7 Soil moisture heterogeneity

There are two options in JULES to introduce sub-gridscale heterogeneity into the soil moisture. One (TOPMODEL) is a more complex scheme that represents this heterogeneity throughout the soil column, including aspects such as a water table and the capability to estimate wetland fractions (Gedney et al., 2004). However, this scheme requires additional topographic ancillary information. The other option (PDM) is a much simpler scheme that does not require as much information. This scheme only considers heterogeneity in the top soil layer and thus can not be used to represent the water table depth or to determine wetland areas. However, it can still be used to increase surface runoff and has been shown to improve subsequent river discharge when fed into a river routing scheme.

4.7.1 TOPMODEL

JULES can optionally use a parameterisation based on TOPMODEL (Beven and Kirkby, 1979). TOPMODEL was initially designed to include a groundwater model within a single catchment where the height of the saturated zone moves up and down and is controlled by the recharge into it and the saturated lateral flow (baseflow) out. As the water table becomes higher, more of the surface area becomes saturated (and vice versa), with the regions of higher topographic index ($\bar{\lambda}_i$) flooding first. Topographic index relates to the upstream area draining into a locality and the local slope, which

Table 6. Default parameter values required by JULES for standard non-vegetation surfaces.

Parameter	Urban	Water	Soil	Ice
Snow-covered albedo	0.40	0.80	0.80	0.80
Snow-free albedo	0.18	0.06	−1.00*	0.75
Canopy capacity ($kg\,m^{-2}$)	0.50	0.00	0.00	0.00
Surface conductance ($m\,s^{-1}$)	0.00	0.00	1×10^{-2}	1×10^{6}
Infiltration enhancement factor	0.10	0.00	0.50	0.00
Roughness length (m)	1.00	3×10^{-4}	3×10^{-4}	1×10^{-4}
Canopy heat capacity ($J\,K^{-1}\,m^{-2}$)	0.28×10^{6}	0.00	0.00	0.00
Fractional "canopy" coverage	1.00	0.00	0.00	0.00

* The snow-free albedo for soil is initialised to −1 to allow it to be set through an ancillary field instead.

is a measure of the potential to flood relative to other regions within the catchment.

This distributed catchment-based model is simplified into a semi-distributed model (Sivapalan et al., 1987) for use in climate models by lumping areas of similar topographic index together from one or potentially more catchments. A gridbox mean water table depth $\bar{z_w}$ is calculated, and the probability distribution function (pdf) of the topographic index within the gridbox is then used to describe the relative frequency of occurrence of the topographic indices. The gridbox fraction of the water table that is above the surface may then be calculated. This enables saturation excess runoff to occur in the model before the gridbox soil moisture is totally saturated. Runoff occurs when water is unable to permeate the fraction of the gridbox surface where the water table is above the surface.

The implementation of this approach in JULES was adapted by Gedney and Cox (2003) and Clark and Gedney (2008). With the TOPMODEL-based approach the free drainage lower boundary condition is replaced by a no flux condition, and sub-surface runoff is represented as a lateral "baseflow", described below. An extra soil layer, with simplified representation of water fluxes, is added beneath the standard soil model as a computationally efficient way in which to track the water table when it is deeper than the standard 3 m soil column. JULES assumes an exponential decrease of K_s in this deeper layer with a decay constant f. The lateral sub-surface runoff, or baseflow, is calculated as

$$R_b = \bar{T}(\bar{z_w})\exp(-\bar{\lambda_i}) \tag{68}$$

where $T(\bar{z_w})$ is the vertical transmissivity from the bottom of the column to the $\bar{z_w}$. This transmissivity is found by summing the contributions from each layer and only considers unfrozen soil water. The transmissivity of each layer is used to partition the total baseflow between soil layers, to give the layer values R_{bn} required in Eq. (56). $\bar{z_w}$ is calculated by assuming that the column soil moisture is in equilibrium (Koster et al., 2000).

The "critical" value of the topographic index at which the water table reaches the surface is found as $\bar{\lambda_{ic}} = \ln(R_{b\,max}/R_b)$, where $R_{b\,max}$ is the baseflow found from Eq. 68 with $\bar{z_w}$=0. The fraction of the gridbox that is saturated (f_{sat}) can be found by integrating the pdf of the topographic index. However, this requires numerical integration if a two-parameter gamma distribution is used for the pdf as in Gedney and Cox (2003). Instead, during the initialisation an exponential distribution is fitted to the results of the gamma distribution, and subsequently f_{sat} is found using

$$f_{sat} = a_s \exp(-c_s f \lambda_{ic}) \tag{69}$$

where a_s and c_s are fitted parameters for each gridbox.

Saturation excess surface runoff (R_{se}) is then calculated as

$$R_{se} = f_{sat} W_0 \tag{70}$$

where W_0 is the rate at which water arrives at the soil surface from precipitation and snowmelt (Eq. 49).

The fraction of the gridbox that is considered to be wetland (i.e., stagnant water) for the purposes of methane emissions (f_{wet}) is defined as that part of the gridbox at which $\lambda_{ic} \le \lambda_i \le \lambda_{i\,max}$ where $\lambda_{i\,max}$ is a global parameter. At locations with larger values of λ_i (water higher above surface) the water is assumed to be flowing and not wetland. Following the procedure for f_{sat}, an exponential relationship is fitted so that f_{wet} can subsequently be calculated as $f_{wet} = a_{wt}\exp(-c_{wt} f \lambda_{ic})$ for parameters a_{wt} and c_{wt}. Gedney and Cox (2003) and Clark and Gedney (2008) showed that simulated runoff was improved by using a TOPMODEL type parameterisation, and that the global pattern of wetland is captured by this model (Gedney and Cox, 2003).

4.7.2 Probability Distribution Model (PDM)

An alternative to TOPMODEL is to calculate saturation excess runoff following the Probability Distributed Model (PDM, Moore, 1985). The distribution of soil storage capacity within a gridbox is modelled by a pdf, and the saturated fraction of the gridbox can be shown to be

$$f_{\rm sat} = 1 - [1 - \theta/\theta_{\rm s}]^{B/(B+1)} \quad (71)$$

where B is a shape parameter. $R_{\rm se}$ is then calculated using Eq. 70. In JULES, B is kept constant across the domain, as is the depth over which W and $W_{\rm max}$ are calculated (typically 1 m). The calculations of infiltration excess and sub-surface runoff are not altered if PDM is selected. Clark and Gedney (2008) showed that the use of PDM improved modelled runoff in mesoscale catchments.

5 Summary

The Joint UK Land Environment Simulator (JULES) is a new community land surface model, based upon the established Met Office Surface Exchange Scheme (MOSES). In addition to representing the exchange of fluxes of heat and moisture between the land surface and the atmosphere (as described here), the model also represents fluxes of carbon and some other gases such as ozone and methane (described in Clark et al., 2011). This enables JULES to be used for many applications, and results in it being a unique land surface model in the fact that it is used in both operational weather forecasting and leading climate change simulations.

Unlike many land surface models, JULES has an explicit representation of the surface energy balance for vegetation, capturing the weaker coupling that exists between the canopy and underlying soil. Other models (e.g., TESSEL, Viterbo and Beljaars, 1995) represent this weaker coupling by adjusting the thermal properties of the top soil layer, but do not have the flexibility of representing radiative, turbulent and conductive exchanges that can be represented in JULES.

Like most other land surface models, JULES uses a tiled land surface scheme to represent heterogeneity in land cover. Many land models have fixed descriptions of the surface types that are designed with specific applications in mind. However, the flexible structure within JULES enables the description of the resolved surface types to be targeted for specific applications. This means, for instance, that there can be a small number of vegetation types for weather forecasting applications where computation cost is critical, but many vegetation types for climate modelling where an accurate representation of the various biomes is important. In addition, JULES introduces elevation bands to the surface types, which is not common in land surface models. The elevated surfaces enable a modified surface energy balance which can be critical for the evolution of snowmelt and sublimation.

Another feature of the snow scheme within JULES is the ability to separate snow held in vegetation canopies and the snow under the canopy, although many other models also make this distinction (e.g., CLASS, CLM). This reduces the spuriously enhanced sublimation of the snow due to an incorrectly increased surface roughness from the tall vegetation components. The new multiple layer snow scheme within JULES also impacts on the timing of snowmelt through the introduction of both solid and liquid water stores. Other land models have a range in the number of snow layers that are modelled, for instance, CLASS uses one layer (Bartlett et al., 2006) whereas CLM uses up to five (Oleson et al., 2010), whilst ISBA has both an implicit snow layer (Douville et al., 1995) or a three-layer snow model (Boone and Etchevers, 2001). However, the majority of snow schemes include both solid and liquid stores within their layers.

JULES has a multilayer approach for both radiation interception and photosynthesis for vegetation. This has been shown to give an improved diurnal cycle for photosynthesis compared to the big leaf approach to scale from leaf to canopy level, that uses only beers law for light interception through the canopy, but is used in some models (such as LPJ, Sitch et al., 2003). Other models do use a multilayer canopy scheme for photosynthesis, but still use beer's law for light interception (e.g., LPJ GUESS, Smith et al., 2001).

There is a selection of three possible options for representing urban surfaces within JULES. All three options have been shown to give a good representation of sensible and latent heat fluxes over urban surface (e.g., in the first international urban model comparison experiment, Grimmond et al., 2010, 2011). The urban surface is integrated into the general framework of the land model, unlike some other models that have to couple an urban model to a separate land model (e.g., ISBA and TEB, Noilhan and Planton, 1989; Masson, 2000).

The heterogeneity of soil moisture can be represented with two methods of varying complexity within JULES. The simple method represents the heterogeneity in the top soil layer only, but can generate increased surface runoff, whereas the more complex scheme has a representation of the mean water table depth. Whilst some land models include the more complex scheme, many do not include soil moisture heterogeneity at all, whilst few have the simpler more computationally efficient method.

The JULES model has been designed with a flexible and modular structure, which means that new elements of science can easily be introduced as new modules into the model. The scientific developments for each module are co-ordinated by an expert in the relevant area of science, ensuring that the model will remain a state of the art land surface model for the research community.

Appendix A

Definitions of symbols

Symbol	Units	Equation	Definition
A	mol CO_2 m^{-2} s^{-1}	8, 11	net photosynthesis uptake
A_m	kg m^{-2}	15	puddling of water on soil surface and interception by leafless vegetation
A_P	mol CO_2 m^{-2} s^{-1}	10, 11	potential leaf level photosynthesis
b		58, 59	Clapp and Hornberger (1978) soil exponent
B_L	kgC m^{-2}	5	leaf biomass
B_m	kgC m^{-2}	15	rate of change of water holding capacity with leaf area index
B_W	kgC m^{-2}	5	woody biomass
C	kg m^{-2}	46, 47, 48	canopy water
C_a	J m^{-3} K^{-1}	53	volumetric heat capacity of the soil
C_c	Pa	8, 9	leaf surface carbon dioxide concentration
C_i	Pa	8, 9	internal leaf carbon dioxide concentration
C_{ice}	J kg^{-1} K^{-1}	22	specific heat capacity of ice
C_k	J m^{-2} K^{-1}	22, 26, 27, 28	areal heat capacity of the k-th snow layer
C_L	J kg^{-1} K^{-1}	5	specific heat capacity of leaves
C_m	kg m^{-2}	15, 46, 48	vegetation canopy water holding capacity
c_p	J kg^{-1} K^{-1}	2, 4	specific heat capacity of air
C_s	J m^{-2} K^{-1}	1, 5	areal heat capacity associated with the surface material
C_W	J kg^{-1} K^{-1}	5	specific heat capacity of wood
C_{water}	J kg^{-1} K^{-1}	22, 55	specific heat capacity of water
C_*	Pa	9	CO_2 compensation point
D	kg kg^{-1}	9	leaf humidity deficit
D_f	m^2 s^{-1}	16	diffusivity of water vapour in air
d_k	m	25	depth of k-th snow layer
d_m	m	38	surface masking snow depth
d_r	m	50	root depth
d_s	m	18, 19, 20, 38, 45	snow depth
d_0	m	39	fresh snow depth
E	kg m^{-2} s^{-1}	1, 3, 6	turbulent moisture flux
E'	kg m^{-2} s^{-1}	56	evapotranspiration
f_{sat}		69, 70, 71	fraction of gridbox with saturated soil
G	W m^{-2}	1, 4, 53	soil heat flux
g	m s^{-2}	21	acceleration due to gravity
G_r	μm^2 s^{-1}	39, 40	snow grain area growth rate
g_s	m s^{-1}	8	leaf level stomatal conductance
g_{soil}	m s^{-1}	7	bare soil surface conductance
H	W m^{-2}	1, 2	turbulent heat flux
h	m	13	height of vegetation canopy
H_k	W m^{-2}	24, 26, 27, 28	conducted heat flux at the bottom of the k-th snow layer
I	kg m^{-2}	16, 17	intercepted snow load
I_0	kg m^{-2}	17	initial intercepted canopy snow load
I_k	kg m^{-2}	22, 34	ice content of k-th snow layer
I_{max}	kg m^{-2}	16, 17	snow interception canopy capacity
J	W m^{-3}	53, 55	vertical advective flux for soil moisture
K	kg m^{-2} s^{-1}	48	surface infiltration rate
k		14	light extinction coefficient
K_e		64, 65	Kersten number
K_h	m s^{-1}	57, 59, 61	hydraulic conductivity
K_{hs}	m s^{-1}	59, 61	hydraulic conductivity for saturated soil
k_l	K^{-1}	37	snow ageing parameter
L	m^2 m^{-2}	14, 15	leaf area index

Symbol	Units	Equation	Definition
L_c	$J\,kg^{-1}$	1	latent heat of condensation of water at 0 °C
L_f	$J\,kg^{-1}$	35, 36	latent heat of fusion
$Lw_{\downarrow}$	$W\,m^{-2}$	1	downward component of the longwave radiation
M_k	kg	21	mass of snow above the middle of the k-th snow layer
n		21	van Genuchten (1980) soil parameter
Q_1	$kg\,kg^{-1}$	3, 6	specific humidity at the reference atmospheric level
$Q_{sat}(T)$	$kg\,kg^{-1}$	3, 6	saturated specific humidity at the temperature T
R	$kg\,m^{-2}\,s^{-1}$	46, 47, 48	Precipitation rate
r	µm	16, 39, 41, 42	snow grain size
r_a	$s\,m^{-1}$	2, 3, 6	aerodynamic resistance
$r_{a_{can}}$	$s\,m^{-1}$	4	aerodynamic resistance between the surface canopy
		4	of vegetation and the underlying soil
R_b	$kg\,m^{-2}\,s^{-1}$	56, 68	lateral runoff
r_e	µm	43	effective snow grain size
R_{es}	$kg\,m^{-2}\,s^{-1}$	70	saturation excess surface runoff
r_k		50, 51	fraction of roots in the k-th soil layer
r_s	$s\,m^{-1}$	3, 6, 16	stomatal or surface moisture resistance
r_0	µm	39, 41, 42	fresh snow grain size
S_f	$kg\,m^{-2}$	17, 39	snowfall amount
S_m	$kg\,m^{-2}\,s^{-2}$	49	snowmelt
$Sw_{\downarrow}$	$W\,m^{-2}$	1	downward component of the solar radiation
t	s	1, 21, 26, 27, 28, 29, 30, 31, 32, 33, 39, 46, 47, 48, 53	time
$\bar{T}$	$m^2\,s^{-1}$	68	vertical transmissivity
T_A	K	2	reference level atmospheric temperature
T_c	K	37	snow albedo threshold temperature
T_F	$kg\,m^{-2}\,s^{-1}$	46, 47, 49	throughfall
T_k	K	21, 24, 26, 27, 28	temperature of the k-th snow layer
T_m	K	21, 34, 35, 40	temperature of the melting point for water
T_s	K	54, 55	soil temperature
T_{s1}	K	4, 27, 28	temperature of the first soil level
T_{sk}	K	53	temperature of the k-th soil level
T_*	K	1, 2, 3, 4, 6, 37, 40	surface temperature
U	$m\,s^{-1}$	16	atmospheric wind speed
W_C	mol CO_2 $m^{-2}\,s^{-1}$	10	rubisco limited rate for photosynthesis
W_E	mol CO_2 $m^{-2}\,s^{-1}$	10	rate of transport of photosynthetic products (for C3 plants), or PEP-Carboxylase limitation for photosynthesis (for C4 plants)
W_L	mol CO_2 $m^{-2}\,s^{-1}$	10	light limited rate for photosynthesis
W_k	$kg\,m^{-2}$	22, 35, 36	water content of k-th snow layer
W_0	$kg\,m^{-2}\,s^{-1}$	49, 70	infiltration rate into the soil
W'	$kg\,m^{-2}\,s^{-1}$	56, 57	vertical flux of soil water
Y	$kg\,m^{-2}\,s^{-1}$	48, 49	surface runoff
z	m	54, 55, 57	soil depth
z_k	m	50, 53	depth of the k-th soil layer
z_t	m	50	total depth of soil column
$\bar{z_w}$	m	68	mean water table depth
z_0	m	13, 45	roughness length for momentum
α		1, 14, 38, 44	surface albedo
α_b		14	bare soil albedo
α_{cds}		37	cold deep snow albedo
α_{nir}		42	diffuse near-infrared snow albedo
α_s		37 38 44	snow albedo

Symbol	Units	Equation	Definition
α_v	m^{-1}	60	van Genuchten (1980) soil parameter
α_{vis}		41	diffuse visible snow albedo
α_0		38 44	snow free albedo
α_∞		14	maximum canopy albedo for dense leaf coverage
α'	$kg\,kg^{-1}\,K^{-1}$	6	$\delta Q_{sat}/\delta T$
β		11, 12, 51, 52	soil moisture factor
γ		24, 27, 28, 32, 33	forward timestep weighting parameter
Δz_i	m	18, 19, 20	thickness of the i-th soil layer
ϵ		1, 4	surface emissivity
ϵ_r		46, 48	fraction of gridcell occupied by convective precipitation
ϵ_s		4	emissivity of the underlying soil surface
ζ		19, 20	snow insulation factor
ν		4, 49	fraction of vegetation
η_0	Pa s	21	compactive viscosity
θ	$m^3\,m^{-3}$	12, 56, 58, 59, 60, 62, 65, 71	soil moisture concentration
θ_1	$m^3\,m^{-3}$	7	soil moisture concentration in the top soil layer
θ_c	$m^3\,m^{-3}$	7, 12	soil moisture concentration at critical point
θ_r	$m^3\,m^{-3}$	60	residual soil moisture concentration
θ_s	$m^3\,m^{-3}$	58, 59, 60, 62, 63, 65, 71	soil moisture concentration at saturation
θ_w	$m^3\,m^{-3}$	12	soil moisture concentration at wilting point
λ	$W\,m^{-1}\,K^{-1}$	18, 62, 64	thermal conductivity
λ_{air}	$W\,m^{-1}\,K^{-1}$	63	thermal conductivity of air
λ_{dry}	$W\,m^{-1}\,K^{-1}$	62, 63, 64, 67	thermal conductivity of dry soil
$\bar{\lambda}_i$	ln(m)	68	topographic index
λ_{ic}	ln(m)	69	"critical" value of topographic index
λ_{ice}	$W\,m^{-1}\,K^{-1}$	63, 66	thermal conductivity of ice
λ_k	$W\,m^{-1}\,K^{-1}$	23, 25	thermal conductivity of k'th snow layer
λ_{snow}	$W\,m^{-1}\,K^{-1}$	18, 20	thermal conductivity of the snow
λ_s	$W\,m^{-1}\,K^{-1}$	62, 63, 64	thermal conductivity for saturated soil
λ_{soil}	$W\,m^{-1}\,K^{-1}$	4, 18, 20	thermal conductivity of the soil
λ_{water}	$W\,m^{-1}\,K^{-1}$	63, 66	thermal conductivity of water
μ		43	cosine of the zenith angle
ν_j		4, 49	fraction of gridbox covered by surface type j
ρ	$kg\,m^{-3}$	2, 3, 4, 6	density of air
ρ_k	$kg\,m^{-3}$	21, 23	density of the k-th snow layer
ρ_i	$kg\,m^{-3}$	16	density of ice
ρ_{water}	$kg\,m^{-3}$	23	density of water
ρ_0	$kg\,m^{-3}$	21	reference snow density
σ	$W\,m^{-2}\,K^{-4}$	1, 4	Stefan Boltzmann constant
Ψ	m	57, 58, 60	soil water suction
Ψ_s	m	58	saturated soil water suction
ω	$m\,m^{-1}$	13	rate of change of roughness length with vegetation canopy height

Supplementary material related to this article is available online at: http://www.geosci-model-dev.net/4/677/2011/gmd-4-677-2011-supplement.pdf.

Acknowledgements. M. Best, M. Pryor, N. Gedney and O. Boucher were supported by the Joint DECC/Defra Met Office Hadley Centre Climate Programme (GA01101).

Edited by: J. C. Hargreaves

References

Bartlett, P. A., MacKay, M. D., and Verseght, D. L.: Modified snow algorithms in the Canadian Land Surface Scheme: Model runs and sensitivity analysis at three boreal forest stands., Atmos. Ocean, 44, 207–222, 2006.

Beljaars, A. C. M. and Holtslag, A. A. M.: Flux parametrisation over land surfaces for atmospheric models, J. Appl. Meteorol., 30, 327–341, 1991.

Best, M. J.: Representing urban areas within operational numerical weather prediction models, Bound.-Lay. Meteorol., 114, 91–109, 2005.

Best, M. J. and Hopwood, W. P.: Modelling the local surface exchange over a grass field site under stable conditions, Q. J. R. Meteorol. Soc., 127, 2033–2052, 2001.

Best, M. J., Beljaars, A., Polcher, J., and Viterbo, P.: A proposed structure for coupling tiled surfaces with the planetary boundary layer, J. Hydrometeorol., 5, 1271–1278, 2004.

Best, M. J., Cox, P. M., and Warrilow, D.: Determining the optimal soil temperature scheme for atmospheric modelling applications, Bound.-Lay. Meteorol., 114, 111–142, 2005.

Best, M. J., Grimmond, C. S. B., and Villani, M. G.: Evaluation of the urban tile in MOSES using surface energy balance observations, Bound.-Lay. Meteorol., 118, 503–525, 2006.

Betts, R. A., Boucher, O., Collins, M., Cox, P. M., Falloon, P., Gedney, N., Hemming, D., Huntingford, C., Jones, C. D., Sexton, D., and Webb, M.: Projected increase in continental runoff due to plant responses to increasing carbon dioxide, Nature, 448, 1037–1041, 2007.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24, 43–69, 1979.

Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M., and Sitch, S.: A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale, Geosci. Model Dev., 4, 255–269, doi:10.5194/gmd-4-255-2011, 2011.

Bohnenstengel, S. I., Evans, S., Clark, P., and Belcher, S. E.: Simulations of the London urban heat island, Q. J. Roy. Meteorol. Soc., 137, doi:10.1002/qj.855, 2011.

Boone, A. and Etchevers, P.: An Intercomparison of Three Snow Schemes of Varying Complexity Coupled to the Same Land Surface Model: Local-Scale Evaluation at an Alpine Site., J. Hydromet., 2, 374–394, 2001.

Bowling, L., Lettenmaier, D., Nijssen, B., Graham, L. P., Clark, D., El Maayar, M., Essery, R., Goers, S., Gusev, Y., Habets, F., van den Hurk, B., Jin, J., Kahan, D., Lohmann, D., Ma, X., Mahanama, S., Mocko, D., Nasonova, O., Niu, G. Y., Samuelsson, P., Shmakin, A., Takata, K., Verseghy, D., Viterbo, P., Xia, Y., Xue, Y., and Yang, Z. L.: Simulation of high-latitude hydrological processes in the Torne-Kalix basin: PILPS Phase 2(e) 1: Experiment description and summary intercomparisons, Glob. Planet. Change, 38, 1–30, 2003.

Brooks, R. H. and A. T. Corey: Hydraulic properties of porous media, Hydrological Papers 3, Colorado State Univ., Fort Collins, 1964.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resour. Res., 14, 601–604, 1978.

Clark, D. B. and Gedney, N.: Representing the effects of of subgrid variability of soil moisture on runoff generation in a land surface model, J. Geophys. Res., 113, D10111, doi:10.1029/2007JD008940, 2008.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description – Part 2: Carbon fluxes and vegetation dynamics, Geosci. Model Dev., 4, 701–722, doi:10.5194/gmd-4-701-2011, 2011.

Collatz, G. J., Ball, J. T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agric. Forest Meteorol., 54, 107–136, 1991.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled photosynthesis-stomatal conductance model for leaves of C_4 plants, Aust. J. Plant Physiol., 19, 519–538, 1992.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils, Water Resour. Res., 20, 682–690, 1984.

Cox, P. M., Huntingford, C., and Harding, R. J.: A canopy conductance and photosynthesis model for use in a GCM land surface scheme, J. Hydrol., 213, 79–94, 1998.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Clim. Dynam., 15, 183–203, 1999.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon–cycle feedbacks in a coupled climate model, Nature, 408, 184–187, 2000.

Cox, P. M., Harris, P., Huntingford, C., Betts, R. A., Collins, M., Jones, C. D., Jupp, T. E., Marengo, J., and Nobre, C.: Increasing risk of Amazonian drought due to decreasing aerosol pollution, Nature, 453, 212–216, 2008.

Cullen, M. J. P.: The unified forecast/climate model, Meteorol. Mag., 122, 81–94, 1993.

Deardorff, J. W.: Efficient prediction of ground surface temperature and moisture with inclusion of a layer of vegetation, J. Geophys. Res., 83, 1889–1903, 1978.

Dharssi, I., Vidale, P. L., Verhoef, A., Macpherson, B., Jones, C., and Best, M.: New soil physical properties implemented in the Unified Model at PS18, Met Office Technical Report 528, Met Office, Exeter, UK, 2009.

Dolman, A. J. and Gregory, D.: The parametrization of rainfall

interception in GCMs, Q. J. R. Meteorol. Soc., 118, 455–467, 1992.

Douville, H., Royer, J.-Y., and Mahfouf, J.-F.: A new snow parameterization for the Meteo-France climate model., Clim. Dynam., 12, 21–35, 1995.

Dyer, A. J.: A review of flux-profile relationships, Bound.-Lay. Meteorol., 7, 363–372, 1974.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., and J. D. Tarpley, J. D.: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, J. Geophys. Res., 108, 163–166, doi:10.1029/2002JD003296, 2003.

Essery, R. and Clark, D.: Developments in the MOSES 2 land-surface model for PILPS 2e, Glob. Planet. Change, 38, 161–164, 2003.

Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M., and Taylor, C. M.: Explicit representation of subgrid heterogeneity in a GCM land surface scheme, J. Hydromet., 4, 530–543, 2003a.

Essery, R. L. H., Pomeroy, J. W., Parvianen, J., and Storck, P.: Sublimation of snow from boreal forests in a climate model, J. Climate, 16, 1855–1864, 2003b.

Essery, R., Rutter, N., Pomeroy, J., Baxter, R., Stahli, M., Gustafsson, D., Barr, A., Bartlett, P., and Elder, K.: SNPWMIP2, An evalutation of froest snow process simulations, Bull. Amer. Met. Soc., 90, 1120–1135, 2009.

Evans, S., Hudson-Smith, A., and Batty, M.: Virtual London and beyond, Cybergeo : European Journal of Geography, A Selection of the best Articles (SAGEO 2005), 359, online on 27 October 2006, modified on 04 July 2007, consulted on 16 June 2011, available at: http://cybergeo.revues.org/2871, last access: 2 August 2011, 2005.

Fisher, J. B., Sitch, S., Malhi, Y., Fisher, R. A., Huntingford, C., and Tan, S. Y.: Carbon cost of plant nitrogen acquisition: A mechanistic, globally applicable model of plant nitrogen uptake, retransocation, and fixation, Global Biogeochem. Cy., 24, GB1014, doi:10.1029/2009GB003621, 2010.

Gedney, N. and Cox, P. M.: The sensitivity of global climate model simulations to the representation of soil moisture heterogeneity, J. Hydromet., 4, 1265–1275, 2003.

Gedney, N., Cox, P. M., and Huntingford, C.: Climate feedback from wetland methane emissions, Geophys. Res. Lett., 31, L20503, doi:10.1029/2004GL020919, 2004.

Gedney, N., Cox, P. M., Betts, R., Boucher, O., Huntingford, C., and Stott, P. A.: Detection of a direct carbon dioxide effect in continental river runoff records, Nature, 439, 835–888, 2006.

Grimmond, C. S. B., Blackett, M., Best, M., Barlow, J., Baik, J.-J., Belcher, S. E., Bohnenstengel, S. I., Calmet, I., Chen, F., Dandou, A., Fortuniak, K., Gouvea, M. L., Hamdi, R., Hendry, M., Kawai, T., Kawamoto, Y., Kondo, H., Krayenhoff, E. S., Lee, S.-H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Pigeon, G., Porson, A., Ryu, Y.-H., Salamanca, F., Shashua-Bar, L., Steeneveld, G.-J., Trombou, M., Voogt, J., Young, D., and Zhang, N.: The international urban energy balance models comparison project: first results from phase 1, J. Appl. Meteorol. Climatol., 49, 1268–1292, doi:10.1175/2010JAMC2354.1, 2010.

Grimmond, C. S. B., Blackett, M., Best, M., Baik, J.-J., Belcher, S. E., Beringer, J., Bohnenstengel, S. I., Calmet, I., Calmet, I., Chen, F., Coutts, A., Dandou, A., Fortuniak, K., Gouvea, M. L., Hamdi, R., Hendry, M., Kanda, M., Kawai, T., Kawamoto, Y., Kondo, H., Krayenhoff, E. S., Lee, S.-H., Loridan, T., Martilli, A., Masson, V., Miao, S., Oleson, K., Ooka, K., Pigeon, G., Porson, A., Ryu, Y.-H., Salamanca, F., Steeneveld, G.-J., Trombou, M., Voogt, J., Young, D. T., and Zhang, N.: Initial results from phase 2 of the international urban energy balance model comparison, Int. J. Climatol., 30, 244–272, doi:10.1002/joc.2227, 2011.

Harman, I. N. and Belcher, S. E.: The surface energy balance and boundary layer over urban street canyons, Q. J. R. Meteorol. Soc., 132, 2749–2768, 2006.

Harman, I. N., Barlow, J., and Belcher, S. E.: Scalar Fluxes from Urban Street Canyons Part II: Model, Bound.-Lay. Meteorol., 113, 387–410, doi:10.1023/B:BOUN.0000045526.07270.a3, 2004.

Jackson, T. L., Feddema, J. J., Oleson, K. W., Bonan, G. B., and Bauer, J. T.: Parameterization for global of urban characteristics climate modeling, Ann. Assoc. Am. Geogr., 100, 848–865, doi:10.1080/00045608.2010.497328, 2010.

Jacobs, C. M. J.: Direct impact of atmopsheric CO_2 enrichment on regional transpiration, Ph.D. thesis, Wageningen Agricultural University, 1994.

Jogireddy, V., Cox, P. M., Huntingford, C., Harding, R. J., and Mercado, L. M.: An improved description of canopy light interception for use in a GCM land-surface scheme: calibration and testing against carbon fluxes at a coniferous forest, Hadley Centre Technical Note 63, Hadley Centre, Met Office, Exeter, UK, 2006.

Johansen, O.: Thermal conductivity of soils, Ph.d. thesis, University of Trondheim, Norway, 1975.

Kojima, K.: Densification of seasonal snow cover, Proc. International Conference on Low Temperature Science, August 14–19, 1966, Sapporo, Japan, 929–952, 1967.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., and Kumar, P.: A catchment-based approach to modeling land surface processes in a general circulation model, J. Geophys. Res., 105, 24809–24822, 2000.

Leuning, R.: A critical appraisal of a combined stomatal-photosynthesis model for C_3 plants, Plant, Cell Environ., 18, 357–364, 1995.

Lynch-Stieglitz, M.: The development and validation of a simple snow model for the GISS GCM, J. Climate, 7, 1842–1855, 1994.

Macdonald, R. W., Griffiths, R. F., and Hall, D.: An improved method for the estimation of surface roughness of obstacle arrays, Atmos. Environ., 32, 1857–1864, 1998.

Mahfouf, J. F. and Noilhan, J.: Comparative-study of various formulations of evaporation from bare soil using insitu data, J. Appl. Meterol., 30, 1354–1365, 1991.

Masson, V.: A physically-based scheme for the urban energy budget in atmospheric models, Bound.-Lay. Meteorol., 41, 1011–1026, 2000.

Mercado, L. M., Huntingford, C., Gash, J. H. C., Cox, P. M., and Jogireddy, V.: Improving the representation of radiative interception and photosynthesis for climate model applications, Tellus B, 59, 553–565, 2007.

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. M.: Impact of changes in diffuse radiation on the global land carbon sink, Nature, 458, 1014–1018, doi:10.1038/nature07949, 2009.

Monin, A. S. and Obukhov, A. M.: Basic regularity in turbulent mixing in the surface layer of the atmosphere, Moscow,

Ak. Nauk, Geof. Inst., 24, 163–187, 1954.
Monsi, M. and Saeki, T.: Ueber den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung fuer die Stoffproduktion, Jpn. J. Bot., 14, 22–52, 1953.
Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrol. Sci. J., 30, 273–297, 1985.
Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, Mon. Weather Rev., 117, 536–549, 1989.
Oleson, K. W., Dai, Y., Bonan, G., Bosilovich, M., Dickinson, R., Dirmeyer, P., Hoffman, F., Houser, P., Levis, S., Niu, G.-Y., Thornton, P., Vertenstein, M., Yang, Z.-L., and Zeng, X.: Technical description of the Community Land Model (CLM), 173 pp., 2004.
Oleson, K. W., Lawrence, D. M., Bonan, G. B., Flanner, M. G., Kluzek, E., Lawrence, P. J., Levis, S., Swenson, S. C., Thornton, P. E., Dai, A., Decker, M., Dickinson, R., Feddema, J., Heald, C. L., Hoffman, F., Lamarque, J. F., Mahowald, N., Niu, G. Y., Qian, T., Randerson, J., Running, S., Sakaguchi, K., Slater, A., Stockli, R., Wang, A., Yang, Z. L., Zeng, X., and Zeng, X.: Technical Description of version 4.0 of the Community Land Model (CLM), 257 pp., 2010.
Pitman, A. J., Liang, Z. L., Cogley, J. G., and Henderson-Sellers, A.: Description of bare essentials of surface transfer for the Bureau of Meteorological Research Centre AGCM, 1991.
Porson, A., Harman, I. N., Bohnenstengel, S. I., and Belcher, S. E.: How many facets are needed to represent the surface energy balance of an urban area?, Bound.-Lay. Meteorol., 132, 107–128, 2009.
Porson, A., Clark, P. A., Harman, I. N., Best, M. J., and Belcher, S. E.: Implementation of a new urban energy budget scheme in the MetUM. Part I. Description and idealized simulations, Q. J. R. Meteorol. Soc., 136, 1514–1529, 2010a.
Porson, A., Clark, P. A., Harman, I. N., Best, M. J., and Belcher, S. E.: Implementation of a new urban energy budget scheme in the MetUM, Part II. Validation against observations and model intercomparison, Q. J. R. Meteorol. Soc., 136, 1530–1542, 2010b.
Schlosser, C. A., Slater, A. G., Pitman, A. J., Vinnikov, K. Y., Henderson-Sellers, A., Speranskaya, N. A., and Mitchell, K.: Simulations of a boreal grassland hydrology at Valdai, Russia: PILPS phase 2(d), Mon. Weather Rev., 128, 301–321, 2000.
Sellers, P. J.: Canopy Reflectance, Photosynthesis, and Transpiration III, A Reanalysis Using Improved Leaf Models and a New Canopy Integration Scheme, Int. J. Remote Sens, 6, 1335–1372, 1985.
Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J. O., Levis, S., Lucht, W., Sykes, M. T., Thonicke, K., and Venevsky, S.: Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model, Glob. Change Biol., 9, 161–185, 2003.
Sitch, S., Cox, P. M., Collins, W. J., and Huntingford, C.: Indirect radiative forcing of climate change through ozone effects on the land-carbon sink, Nature, 448, 791–794, 2007.
Sivapalan, M., Wood, E. F., and Beven, K. J.: On hydrologic similarity. 2. A scaled model of storm runoff production. 3. A dimensionless flood frequency model using a generalised geomorphologic unit hydrograph and partial area runoff generation, Water Resour. Res., 26, 43–58, 1987.
Smith, B., Prentice, I. C., and Sykes, M. T.: Representation of vegetation dynamics in the modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space., Global Ecol. Biogeogr., 10, 621–637, doi:10.1046/j.1466-822X.2001.t01-1-00256.x, 2001.
Sokolov, A. P., Kicklighter, D. W., Melillo, J. M., Felzer, B. S., Schlosser, C. A., and Cronin, T. W.: Consequences of considering carbon-nitrogen interactions on the feedbacks between climate and the terrestrial carbon cycle, J. Climate, 21, 3776–3796, doi:10.1175/2008JCLI2038.1, 2008.
Storck, P., Lettenmaier, D. P., and Bolton, S. M.: Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States, Water Resour. Res., 38, 1223, doi:10.1029/2002WR001281, 2002.
Taylor, C. M. and Clark, D. B.: The diurnal cycle and African easterly waves: A land surface perspective, Q. J. R. Meteorol. Soc., 127, 845–867, 2001.
Thornton, P. E., Lamarque, J. F., Rosenbloom, N. A., and Mahowald, N. M.: Influence of carbon-nitrogen cycle coupling on land model response to CO2 fertilization and climate variability, Global Biogeochem. Cy., 21, GB4018, doi:10.1029/2006GB002868, 2007.
Thornton, P. E., Doney, S. C., Lindsay, K., Moore, J. K., Mahowald, N., Randerson, J. T., Fung, I., Lamarque, J. F., Feddema, J. J., and Lee, Y. H.: Carbon-nitrogen interactions regulate climate-carbon cycle feedbacks: results from an atmosphere-ocean general circulation model, Biogeosciences, 6, 2099–2120, doi:10.5194/bg-6-2099-2009, 2009.
van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils., Soil Sci. Soc. America J., 44, 892–898, 1980.
Verseghy, D. L.: CLASS – a Canadian land surface scheme for GSMS, I. Soil model, Int. J. Climatol., 11, 347–370, 1991.
Verseghy, D. L., Lazare, M., and McFarlane, N.: CLASS – A Canadian land surface scheme for GSMS, II. vegetation model and coupled runs, Int. J. Climatol., 13, 111–133, 1993.
Viterbo, P. and Beljaars, A. C. M.: An Improved Land-Surface Parameterization Scheme in the Ecmwf Model and Its Validation, J. Climate, 8, 2716–2748, 1995.
Wiscombe, W. J. and Warren, S. G.: A model for the spectral albedo of snow, 1. Pure snow, J. Atmos. Science, 37, 2712–2733, 1980.
Yen, Y.-C.: Review of thermal properties of snow, ice and sea ice, 27 pp., 1981.
Zaehle, S., Friend, A. D., Friedlingstein, P., Dentener, F., Peylin, P., and Schulz, M.: Carbon and nitrogen cycle dynamics in the O-CN land surface model: 2. Role of the nitrogen cycle in the historical terrestrial carbon balance, Global Biogeochem. Cy., 24, GB1006, doi:10.1029/2009GB003522, 2010.