Catchment-level effects on river habitats
a spatial data-science study of rivers in England and Wales

Miles, Eleanore

Awarding institution:
King's College London

CATCHMENT-LEVEL EFFECTS ON RIVER HABITATS

A spatial data-science study of rivers in England and Wales

Eleanore Lucy Heasley Miles

Thesis submitted to King's College London
for the degree of Doctor of Philosophy

Department of Geography

King's College London

Abstract

An appreciation of catchment-level effects (i.e. the impacts from catchment characteristics such as morphometry, climate, geology, land cover and the drainage network) on river reaches is often seen as the gold standard in river management. Yet quantifying catchment-level effects remains a complex research area. With branches in geomorphology, hydrology, ecology and applied river management, studies are often restricted to individual sites or catchments and only frequently consider a subset of possible catchment controls, primarily anthropogenic drivers. This PhD provides a more holistic, possibly transferrable methodology and interpretation, considering multiple catchment-level effects and focusing on an often-overlooked component of the catchment: the topology of the river network.

To achieve this, a broad-scale approach is adopted utilising a monitoring dataset collected for regulatory compliance (the River Habitat Survey) and adapting it for scientific enquiry. When paired with GIS-derived catchment controls and data-science techniques, this monitoring dataset enables national-level enquiry into catchment-level effects on the type and diversity of physical habitats in river reaches in England. Here, catchment-level effects are quantified via: (i) the production of a national waterbody typology combining multiple catchment-level effects using machine learning techniques; and (ii) the adaptation of flood estimation metrics to reflect network topological structure. Statistical analysis of the monitoring dataset shows that both the waterbody typology and network topology have functional applicability to physical habitats.

This PhD not only aims to provide new ways of quantifying catchment-level effects but also aims to improve our understanding of their impacts. To accomplish this, controls from multiple spatial hierarchical levels in the river system – from catchment to reach – are combined using a data-science approach to explain the controls on physical habitat type and diversity in river reaches in England. The results show that there are broad patterns in physical habitats from upland to lowland catchments, and upstream to downstream within catchments. These results are consistent with earlier analysis of the River Habitat Survey dataset. However, there remains much variation, only some of which is explained by the influence of network topology. This national-level assessment of catchment-level effects demonstrates the importance of more holistic and strategic thinking in river management. The transferable methodologies enable river managers to better spatially target areas for management or conservation within catchments and compare sites with similar catchment-level effects.

Acknowledgements

This research was conducted in the Department of Geography, King's College London and was funded by the Natural Environment Research Council (NERC).

I thank my first supervisor Dr James Millington for his continual support and optimism throughout this PhD process. From discussions around statistics to proof reading chapter drafts and manuscripts, James' enthusiasm generated my interest in creative approaches to data analysis. I also thank my second supervisor Dr Michael Chadwick for his ongoing encouragement and positivity throughout my time at King's. I would also like to thank Prof. Nick Clifford at University of Loughborough for many interesting and informative discussions that have developed my understanding of the river system and his continued supervision that has helped shape my PhD.

My PhD experience has been enjoyable, challenging and inspiring and I would like to thank all those at King's who have provided academic support as well as much needed tea breaks and pub trips over the past four years. Special thanks to Rory Walshe, Fenner Holman, Lara Langston, Mark de Jong, Anna Turbelin, Hannah Nguyen, Mary Langsdale, Tom Dowling, Dan Mills and Anna Lavelle.

This work would not have been possible without the unwavering support and encouragement of my husband Ryan Miles, and my family and friends who reminded me there was a world outside of my PhD. My final thank you goes to my parents who sparked my interest in the natural world from an early age and continue to inspire me every day. I dedicate this thesis to my parents, Brian and Shelagh Heasley.

Table of Contents

List of Figures

List of Tables

CHAPTER 1

Introduction and Aims

Rivers are 'integrators' of the characteristics in their catchments (Dovers and Day, 1988) so that impacts from upstream flow downstream to river reaches. This thesis investigates how catchment characteristics influence habitat features within reaches. The importance of catchment controls on river reaches is increasingly highlighted in river science and management (Defra, 2013), particularly effects on river habitats which are often the focus of river management and restoration plans at specific sites (Newson and Newson, 2000). However, catchment controls are rarely captured holistically or compared between multiple catchments. Therefore, it is critical to explain how catchment influence varies spatially at a national level for better strategic management and understanding of reach responses to catchment-level controls.

Many aspects of the ways in which catchment-level effects are functionally observed are explored in this thesis, such as catchment topography, geology, climate, land cover and the river network. Emphasis is placed the structure of the river network, or network topology, as an integrator of catchment-level effects. Until now network topology has been underrepresented in studies of catchment-level effects yet it plays a role in catchment functioning by connecting upstream reaches to downstream reaches (Benda *et al.*, 2004b; Rice, 2017; Vannote *et al.*, 1980). This thesis will explore in detail the effect of network topology on reach-level habitats and combine the relationships identified with other aspects of the catchment to gain a more holistic understanding of catchment-level effects. Three broad aims of this thesis are described below and in **Figure 1.1**:

Aim 1: Integrate a theoretical understanding of the river network with regulatory monitoring data to identify catchment-level effects on river habitat.

Aim 2: To improve understanding of the roles network and catchment structure can play in assessing habitat distribution within and between catchments.

Aim 3: Apply knowledge of catchment-level effects to explain national patterns of river habitats and explore management applications.

The thesis will explore these aims using an extensive approach to research design which uses a large number of empirical samples to identify and quantify relationships between entities (Sayer, 1992). Therefore, the focus of this thesis is on general pattern and form of relationships rather than mechanisms or individual case studies. The rationale for an extensive approach in this thesis is clear; to move beyond the site-specific examples of catchment-level effects towards a more holistic understanding of patterns at a national level. To this end, this thesis employs the use of broad-scale monitoring datasets collected for regulatory compliance and adapts them for use in science. Broad-scale datasets are utilised in combination with data-science techniques to explore how the characteristics of the catchment and network impact habitats in river reaches across England. The results of this thesis help fill scientific gaps in knowledge and may help inform management practices to encourage a strategic approach to catchment and river management.

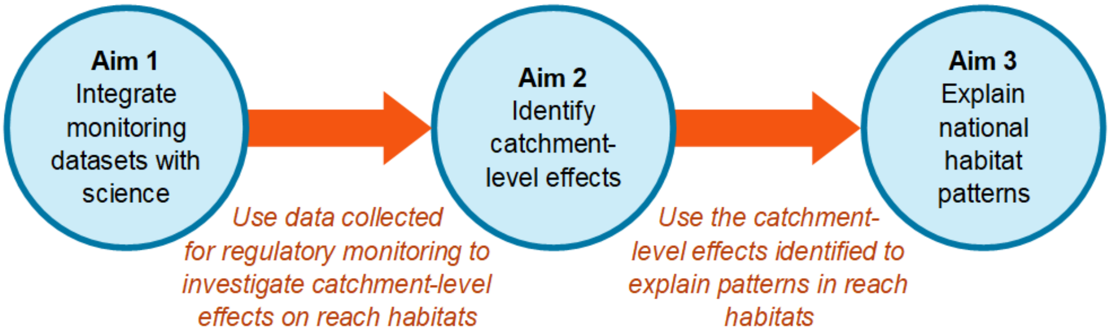


Figure 1.1. *Diagram summarising the connectivity between the thesis aims. How the aims will be delivered is described below each arrow.*

The aims of the thesis are connected (**Figure 1.1**); broad-scale monitoring data are adapted for scientific study and are used to explore catchment-level effects on river habitats. The relationships identified with different catchment characteristics are then combined in a multivariate model to explain habitat patterns across England.

The aims are delivered through chapter specific objectives that are summarised at the end of Chapter 2 after the motivations for the study and gaps in research are described in more detail. The objectives are repeated in the introduction of each data chapter.

1.1 THESIS STRUCTURE

The thesis is structured as described below and the spatial level investigated by each data chapter is presented as a schematic in **Figure 1.2**.

Chapter 2 provides a literature review to highlight the key themes and theories that form the basis of the thesis. It discusses the hierarchy of spatial levels in the river system (Frissell *et al.*, 1986) and how this has influenced river management strategy (Beechie *et al.*, 2010). An evidence review of large data-science studies is presented to identify which catchment characteristics have been used to explain patterns of river features previously, in addition to common practices and research gaps. Finally, the opportunities and limitations of using broad-scale datasets in scientific studies is discussed and a detailed description of the key broad-scale dataset used in this thesis - the River Habitat Survey (RHS) - is provided.

Chapter 3 is the first data analysis chapter and is conducted at the ***catchment level,*** combining multiple characteristics of sub-catchments into a single typology. Appreciation of multiple characteristics in river management typologies is of recognised importance (Downs and Gregory, 2004) but frequently characteristics are considered in isolation, without considering associations between characteristics. This typology classifies the types of sub-catchments (i.e. waterbodies) in England and Wales based on a range of catchment characteristics using a machine learning technique. The waterbody typology is validated against physical habitat features from the RHS dataset (**Figure 1.2**). The chapter also investigates the range of waterbody types within catchments to explore catchment complexity. This chapter is contrary to previous applied classifications which have focused on reach-level characteristics rather than the wider catchment.

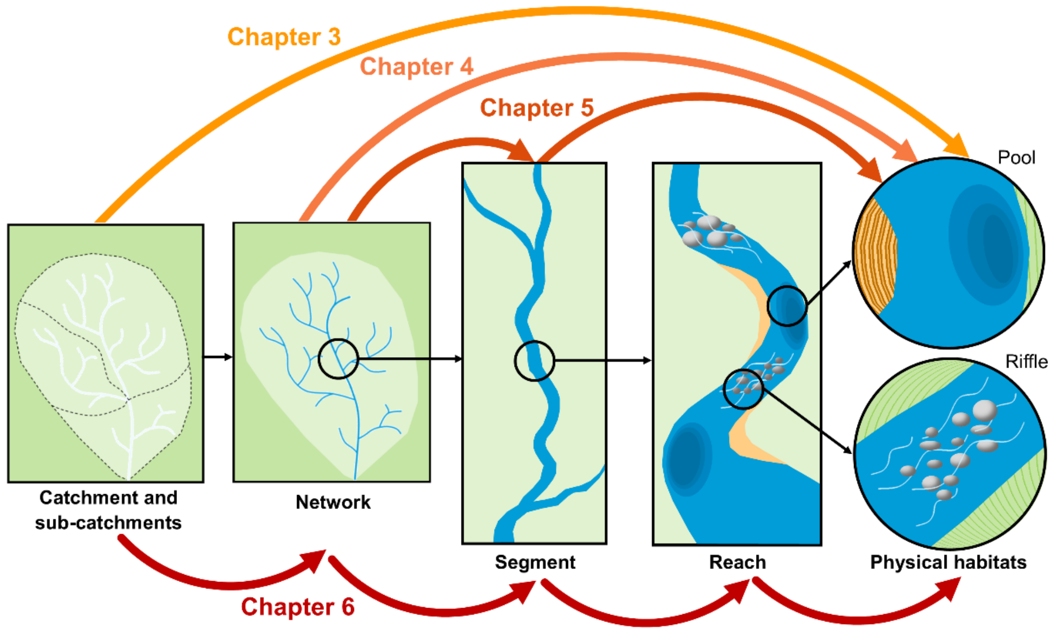


Figure 1.2. *Conceptual model of the thesis structure overlain on a multi-level hierarchical river system that includes the river network and confluences. Spatial levels addressed by each chapter are indicated with arrows. Further details on hierarchical river systems is provided in Chapter 2.1.1.*

Chapter 4 focuses on the catchment characteristic network topology at the ***network level***, a key mechanistic component of the system linking upstream catchment hillslope processes to downstream reaches (Tetzlaff *et al.*, 2007). Within this chapter, simple metrics that quantify network topology and highlight topological variation within catchments are developed that can easily be extracted for any catchment using Geographical Information Systems (GIS). The impact of network topology on the spatial arrangement of physical habitat river features (**Figure 1.2**) is then assessed for four catchments in England using correlation.

Chapter 5 addresses questions posed by the results of Chapter 4. It explores why network topology has different effects on physical habitat features in different catchments and whether the impact of individual confluences in the river network are influencing this difference (Benda *et al.*, 2004b; Rice, 2017). The RHS dataset is used in this chapter to identify which confluences change physical habitat features and further analysis identifies if confluences at the ***segment level*** are concentration points at the ***network level*** for catchment-level effects on physical habitats (**Figure 1.2**).

Chapter 6 combines the waterbody typology (Chapter 3), network metrics (Chapter 4), confluence effects (Chapter 5) and reach characteristics in a multivariate model to explain physical habitat patterns based on combined catchment-level effects (**Figure 1.2**). This is achieved by using a boosted regression tree technique to explore the hierarchical structure and interactions between different catchment-level effects at multiple spatial levels.

Chapter 7 provides a summary of the findings of the thesis, describes the conclusions relating to the overall aims and proposes opportunities for further work.

Research context, design and datasets

This chapter provides an overview of how catchment-level effects are currently included in applied river science and management (Section 2.1.1) and of the concept of physical habitats (Section 2.1.2) – the key control and response variables in this thesis. A brief evidence review is presented that outlines the breadth of the current literature base linking catchment-level effects to physical habitat responses (Section 2.1.3). The evidence review is used to identify gaps in knowledge and place this work within the context of other research in the field.

The data-science methodological approach adopted by the thesis is described along with the strengths and limitations associated with broad-scale data (Section 2.2.1). The key dataset used in this research – the River Habitat Survey – is described and critiqued (Section 2.2.2). Finally, the objectives of this thesis are outlined with reference to the contextual literature (Section 2.3). Specific literature pertaining to each objective is discussed in the relevant chapter.

2.1 RESEARCH CONTEXT

2.1.1 Catchment-level thinking: exploring the catchment-level effect

The catchment is an area of land bounded by topography where water and sediment are transported from the land to the river channel and flow downstream to reaches via the river network. In this thesis, the functional connectivity between the catchment and its reaches is termed the ***catchment-level effect*** and is explored primarily from a hydromorphological perspective. However, the fields of geomorphology, hydrology, ecology and applied river science are interwoven, exhibited by several terms such as ecohydrology and eco-hydromorphology (Vaughan *et al.*, 2009), so this thesis draws on research from multiple disciplines.

Catchment-level effects are of recognised importance. For example, the catchment itself is the fundamental unit of geomorphology and hydrology (Chorley, 1969), and the natural scale to manage the river system through catchment management plans (Defra, 2013). There have been recurrent calls for methods that effectively capture catchment-level effects to encourage integrated catchment management (Beechie *et al.*, 2010; Downs and Gregory, 2004; England and Gurnell, 2016) but there are few holistic examples. This is because of the elusive nature of catchment-level effects and the complexities of scale, connectivity and interaction in the river system.

There are numerous processes which influence the catchment-level effect on river reaches and the key theories are outlined in Section 2.1.1.1. It is critical to understand how catchment-level effects impact river reaches, as individual reaches are the focus of management interventions (see Section 2.1.1.2 for full description of why catchment-level effects are important for river management).

2.1.1.1 Key patterns and process within catchments

The catchment-level effect is a nebulous paradigm due to the complexities of interactions and processes occurring across space and time. This thesis focuses on the spatio-functional dimensions of this cascade due to the broad-scale approach adopted (see Section 2.2 for details of the methodology).

Catchment systems are frequently conceptualised as a cascade of multiple spatial and temporal levels but Frissell *et al.*'s (1986) hydromorphological framework is the most comprehensive (Gurnell *et al.*, 2016). In this framework, levels with a smaller spatial extent are nested within levels with successively larger extents in a hierarchy up to and including the catchment (**Figure 2.1**).

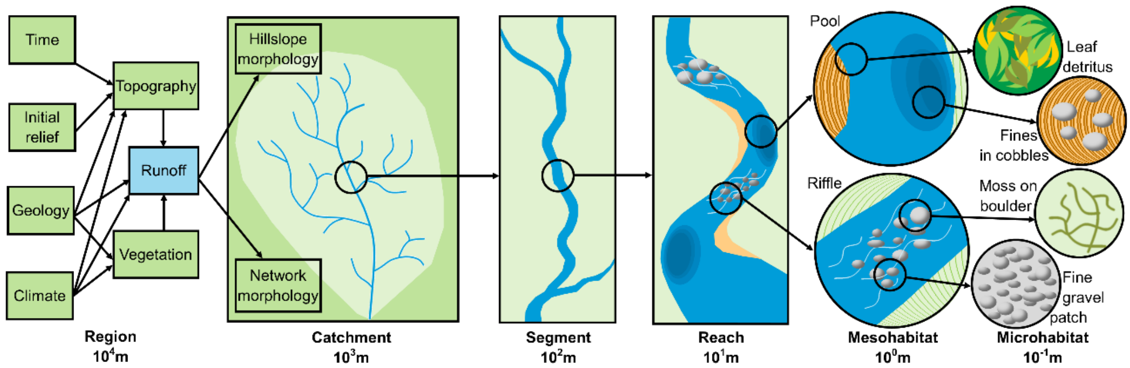


Figure 2.1. *Diagram of the spatial levels in the river system. Hierarchical structure and associated habitats shown with an approximate linear spatial scale* (Frissell *et al.*, 1986). *Regional characteristics influence catchment level that feeds through the hierarchical structure* (Schumm and Lichty, 1965).

Regional controls – such as geology, climate and initial relief – over time determine the boundary conditions of catchment structure and function, including the development of catchment topography and vegetation (Schumm and Lichty, 1965). Regional catchment properties drive water and sediment runoff to alter drainage network and hillslope morphology (Schumm and Lichty, 1965). This controls hydrological and sediment regimes which shape channel morphology through effective discharges (Wolman and Miller, 1960) and thus create physical habitats for instream ecology (Bunn and Arthington, 2002).

Frissell *et al.*'s (1986) framework suggests an interdependency between levels so processes occurring at the catchment level successively alter all smaller levels and vice versa (Sear, 1996) within the regional boundary conditions. The regional level was not included in Frissell *et al.*'s (1986) hierarchy but is incorporated in the schematic in **Figure 2.1** because of its influence on catchment boundary conditions. It has also recently been included in a hierarchical framework designed to aid sustainable river management (England and Gurnell, 2016; Gurnell *et al.*, 2016).

This hydromorphological framework is also adopted by the field of ecology where landscape ecology concepts of scale and patch dynamics are applied to river systems such as riverscapes (Fausch *et al.*, 2002; Wiens, 2002). It has also been adapted to management, highlighting river restoration options that deliver multiple ecosystem services from the reach-level, such as re-meandering, to weir removal across the entire river network (Gilvear *et al.*, 2013).

Along with the hierarchical framework, the other key concept that has developed in river research is the idea of differences between upstream and downstream reaches (**Figure 2.2**), or the longitudinal dimension of the river (Petts and Amoros, 1996). The fluvial system is highly directional, with water and sediment flowing downstream so that river reaches become 'receivers' of influences from upstream (Dudgeon *et al.*, 2006). This concept becomes interwoven with the hierarchical framework, as the arrangement of smaller spatial units within larger spatial units is partly determined by downstream flows.

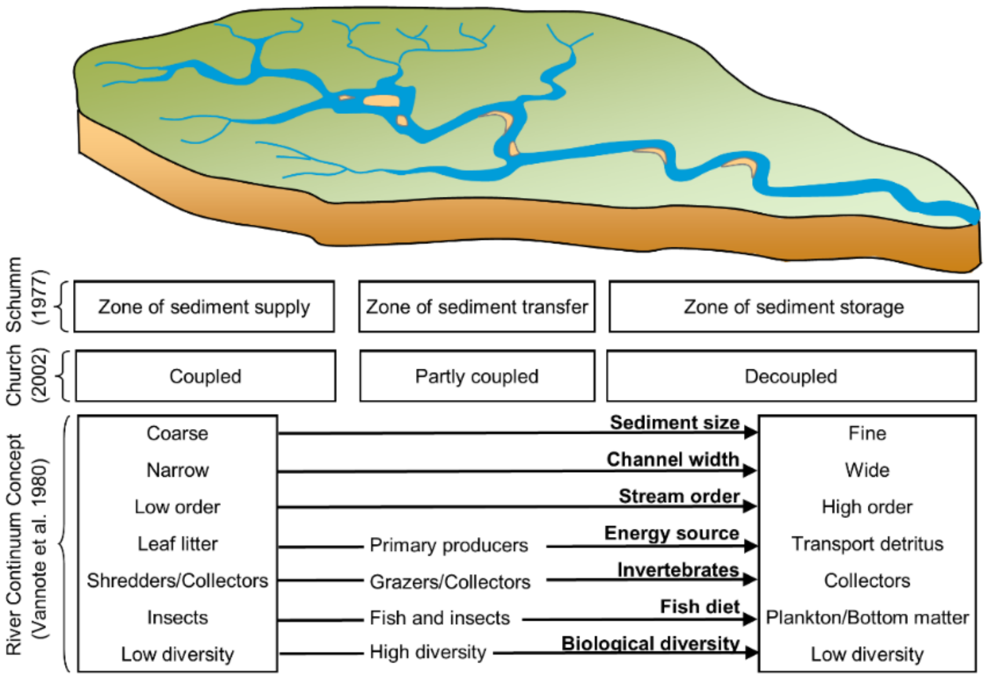


Figure 2.2. *Summary of longitudinal process domains and gradients identified by Schumm (1977), Church (2002) and Vannote et al. (1980).*

Longitudinal patterns in channel features have been observed in geomorphic studies. These geomorphic changes are due to the changing balance between sediment transport capacity and sediment supply downstream. Lane (1955) represents this balance as

$$Q \cdot S \propto Q_s \cdot D_{50}$$

(Equation 2.1)

where Q is discharge, S is channel slope, Q_s is sediment supply and D_{50} is median diameter of sediment supplied. Imbalances between transport capacity ($Q \cdot S$) and sediment supply ($Q_s \cdot D_{50}$) lead to erosion (where $Q \cdot S > Q_s \cdot D_{50}$) or deposition (where $Q \cdot S < Q_s \cdot D_{50}$) resulting in changes to channel width, depth and sinuosity (Schumm, 1977). Studies show increases in width, depth and velocity (Leopold and Maddock, 1953); and decreases in braiding and incision downstream (Piegay *et al.*, 2000). Bedload sediment size is also shown to decrease downstream, known as sediment fining (Ferguson *et al.*, 1996; Werritty, 1992).

The change in geomorphic processes from upstream to downstream has been conceptualised as three process domains (**Figure 2.2**). Schumm's (1977) classification describes: (i) upstream reaches as zones of sediment supply where there is high sediment availability, erosion and transport capacity; (ii) mid reaches as zones of sediment transfer where sediment is repeatedly deposited and eroded; and (iii) downstream reaches as zones of sediment storage where sediment is deposited on the floodplain. This is complemented by Church's (2002) classification where upstream reaches are highly coupled to hillslope processes because they are close to sediment sources and are responsive to rain events. In contrast, downstream reaches have wide floodplains so are decoupled from hillslope processes (**Figure 2.2**).

These geomorphic classifications are supplemented by ecological theories, for example, the Network Position Hypothesis predicts that biotic communities in upstream reaches are more regulated by environmental processes than their downstream counterparts (Schmera *et al.*, 2018). However, perhaps the most influential theory is the River Continuum Concept (RCC; Vannote *et al.*, 1980; **Figure 2.2**). The RCC describes a continuous gradient in physical processes and resulting biotic adjustments in the loading, transport, utilization, and storage of organic matter which alter the ecological communities along the length of a river. Both the geomorphic and ecological classifications (**Figure 2.2**) highlight mid-reaches as the most heterogeneous but also the most sensitive to changes in the catchment (Church, 2002).

While longitudinal patterns in morphology and ecology may occur broadly, there are flaws in the continuum approaches shown in **Figure 2.2**. For example, the RCC is only valid for low-relief watersheds with constant climate and geology and cannot be generalised globally (Montgomery, 1999; Townsend, 1989). A key critique is that trends in instream variables are more complex than a simple longitudinal continuum and lateral inputs from tributaries and other sources cause perturbations that are not considered by the RCC (Petts and Amoros, 1996). They are, however, included in theories that highlight the structure of the network. For example, the Link Discontinuity Concept describes how downstream sediment fining is disrupted by coarse lateral inputs creating a saw-tooth downstream pattern (Rice *et al.*, 2001), and the Network Dynamic Hypothesis predicts how the type and distribution of tributaries can have a marked impact on geomorphic heterogeneity (Benda *et al.*, 2004b). Others suggest that there is no continuum to be perturbed by lateral inputs as rivers are zonal rather than clinal, composed of patches in the landscape (Townsend, 1989).

These concepts highlight the complexities surrounding the catchment-level effect on river reaches. Understanding of these processes and their interactions is limited but is critical to inform sustainable river management practices at a national level as discussed below.

2.1.1.2 Catchment-level effects in river management

There have been numerous and longstanding calls to recognise the catchment as the fundamental unit of management, and catchment management plans have been developed (Defra, 2013), but in practice holistic management is difficult to achieve. For example, Bannister *et al.*'s (2005) review of catchment scale river restoration in the UK (defined as a project that takes into account the main processes and restraints that operate at the catchment-level) identified that no project focused on catchment scale river restoration as its primary driver and that no truly integrated catchment scale river restoration project existed at the time of the review. Similarly, Gilvear *et al.*'s (2012) results from interviews with key river restoration stakeholders in Scotland found instances of catchment level projects, yet the number is limited compared to overall growth in restoration projects.

This is not to say that the motivation for catchment scale management is not present, with widespread acceptance by a range of stakeholders (including local authorities, consultancies, trusts and community organisations) that the catchment influences river reaches and should be considered in river restoration (Gilvear *et al.*, 2012; Gurnell *et al.*, 2016). However, there are barriers to catchment level restoration. For example, funding is frequently only targeted at a single driver, often flood risk, which constrains the location and scale of restoration projects. Stakeholders also stated that government bodies should coordinate catchment-level and national restoration strategy (Gilvear *et al.*, 2012) as catchment-level restoration involves managing multiple stakeholders that often makes it complex and unattainable (Smith, 2015). Integrated catchment management structures are in place to deal with stakeholder relationships and coordinate river management within catchment boundaries rather than administration boundaries, such as the catchment based approach in England (Defra, 2013). However, integrated management may be more politically than technically focused (Campbell, 2016) and is often targeted towards flood control and water supply for industrial, agricultural and domestic uses, rather than maintaining the integrity of the fluvial system (Downs *et al.*, 1991). This highlights that integrated catchment management is only consistently being applied to certain river management applications and does not consider fundamental catchment processes. Therefore, it is not truly integrated.

The discrepancy between appreciating the need for strategic catchment-level assessment and funding limitations is exacerbated by a scarcity of practical tools available to end users that can be cheaply and quickly applied at the catchment level (Parker *et al.*, 2015). Frameworks and datasets exist that could help support the wider inclusion of catchment-level controls in river restoration. For example, the REFORM framework that is targeted at river managers and encourages catchment-level thinking (England and Gurnell, 2016; Gurnell *et al.*, 2016). Databases of previous projects are also available, such as the UK River Restoration Centre's National River Restoration Inventory (NRRI) and EU RiverWiki (restorerivers.eu) used by Smith *et al.* (2014a) for national level assessment of river restoration projects. However, tools that identify catchment-level effects are often developed for a specific project (Bannister *et al.*, 2005) and there are few tools available to easily identify catchment-level effects consistently between catchments to allow for wider strategic management planning.

Therefore, in reality river managers often focus on individual reaches, where monitoring and individual interventions are implemented, when making decisions about management (Bannister *et al.*, 2005; Beechie *et al.*, 2010), with little consideration of the whole river network or catchment-level effects (Harper and Everard, 1998; Kuemmerlen *et al.*, 2019). Lack of consideration of wider catchment influences can have implications for the success and sustainability of interventions. A frequent cause of restoration project failure is attributed to anthropogenic pressures upstream (e.g. urbanisation, agriculture, deforestation, flow regulation and water extraction) that propagate to downstream reaches, negating positive interventions (Cockburn *et al.*, 2015; Lorenz and Feld, 2013;

Palmer *et al.*, 2010). This is because restoration is often conducted opportunistically rather than strategically as highlighted by a review of river restoration trends in the UK (Smith *et al.,* 2014a). Smith *et al.* (2014a) shows that river restoration frequently occurs in urban, lowland or protected sites but higher energy channels, that have greater recovery potential (Downs and Gregory, 2004), are less frequently restored. This is because project sites are usually selected based on land availability even if the site is suboptimal (Palmer, 2009) and funding is targeted at areas in the public interest (e.g. in urban areas and protected systems). Smith *et al.*'s (2014a) assessment showed no changes in the types of catchments where restoration projects were located over a 35 to 45year period indicating that lessons are not being learnt from previous projects. Therefore, care must be taken to ensure the suitability and sustainability of restoration designs in relation to catchment-level effects (Beechie *et al.*, 2010) when effects are not considered during site selection, or when the development of best practice is limited.

In addition, catchment-level effects are also not frequently considered in the science that forms the evidence base for widely implemented river management approaches. For example, many nationally applied methods are based on results of a few experiments at smaller spatial scales (Calder and Aylward, 2006; Newson, 2010). A holistic method for targeted river management requires research at the catchment-level for managers (and regionally, nationally or globally for policy-makers) to highlight how the results apply to wider spatial and temporal scales and under contrasting catchment conditions (Clifford, 2002; Moran *et al.*, 2008; Tetzlaff *et al.*, 2007; Wiens, 2002). This is important to identify whether responses to management interventions differ across all catchments or whether there are systematic differences that reflect catchment-level effects.

Contextualising individual reaches within the wider catchment would be beneficial to: (i) assess the controls that will influence the reach to ensure the intervention strategy is appropriate; (ii) generalise the outcome of interventions to reaches with similar catchment-level effects to learn from similar projects; and (iii) prioritise reaches where natural recovery is more likely and interventions are less likely to be destabilised by adverse upstream conditions (Clarke *et al.*, 2003). There is the need for practical tools to quickly assess catchment-level effects so they may be included in river management decisions to allow in-channel projects to work with natural processes and be sustainable (Sear, 1994). This is because the boundary conditions of the catchment (**Figure 2.1**) and the position within the network (**Figure 2.2**) will determine scale and location of interventions that will have the most effect. This is especially necessary for projects where funding to consider catchment-level effects is limited so all projects can move towards wider catchment integration.

2.1.2 Physical habitats and biotopes

Physical habitats have been the focus of monitoring surveys (Raven *et al.*, 1996), management (Harper and Everard, 1998) and restoration (Kemp *et al.*, 1999). This is because physical habitats, the abiotic conditions needed for aquatic biota to survive (e.g. velocity, depth and substrate), are often measured at the sub-reach meso-level (~10^0m; **Figure 2.1**). The meso spatial level is fine enough to capture sophisticated scientific detail but coarse enough to be measured from the bankside so is practical for surveys (Kershner and Snider, 1992; Newson and Newson, 2000). This thesis focuses on a bottom-up approach to habitat assessment, inferring ecology from hydromorphology, rather than a top-down approach, inferring physical habitat conditions from ecology (Harper *et al.*, 1992; Newson and Newson, 2000).

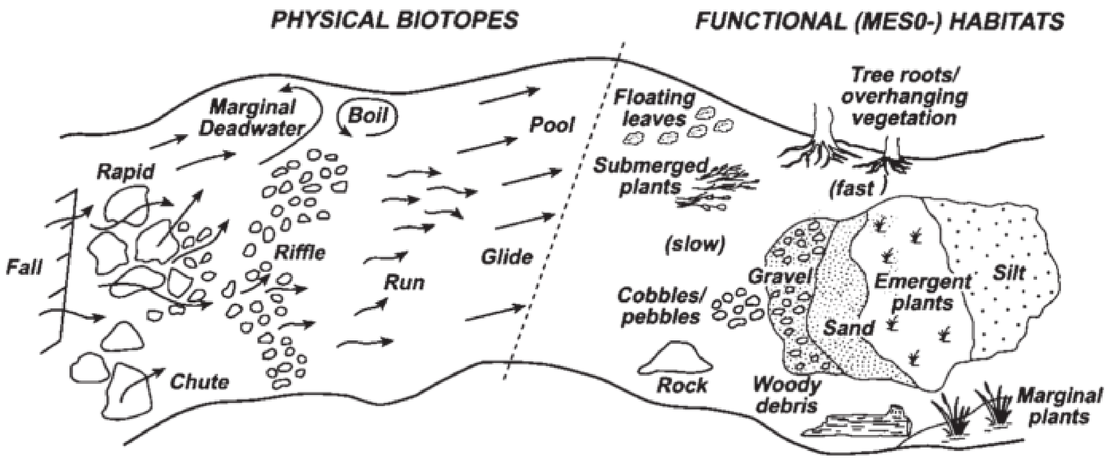


Figure 2.3. *Comparison of biotopes (left) and functional habitats (right) in a hypothetical sub-reach. Reprinted Figure 2 from* Newson and Newson (2000, p.200).

Physical habitat units have been categorised by different methods such as 'functional habitats' (Harper *et al.*, 1995) and 'biotopes' (Jowett, 1993; Padmore, 1997; Rowntree, 1996) (**Figure 2.3**). To briefly summarise, functional habitats represent the physical substrate that biota inhabit consisting of biotic (macrophyte type) and abiotic elements (bed sediment size) (Harper *et al.*, 1992). Functional habitats have been linked to physical habitat parameters such as velocity and depth (Kemp *et al.*, 1999; Pardo and Armitage, 1997). On the other hand, biotopes measure the flow hydraulics of the water column through visual assessment of the surface flow type. Flow types are related to bed morphology (Zavadil *et al.*, 2012), depth, velocity, shear stress and Froude number (an index of velocity and depth) (Jowett, 1993; Newson *et al.*, 1998b; Padmore, 1998; Rowntree and Wadeson, 1998). A biotope is said to reflect the habitat of an entire biological community (Udvardy, 1959) and different macroinvertebrate assemblages have been identified in different biotopes (Hill *et al.*, 2008; Large and Heritage, 2007; Newson *et al.*, 1998b; Reid and Thoms, 2008).

There are strengths and weakness of the physical habitat units displayed in **Figure 2.3**. Biotopes are best suited to upland streams but are influenced by channel vegetation and have overlapping physical conditions (Clifford *et al.*, 2006; Newson *et al.*, 1998b; Wadeson and Rowntree, 1998). They can be measured at high flows but vary according to discharge (Padmore, 1997). Functional habitats are better suited for lowland, vegetated channels but are not suited for high flow conditions as the bed may not be visible (Newson and Newson 2000). The strengths and weaknesses of both methods of habitat derivation complement each other.

Biotopes and functional habitats are related (Newson *et al.*, 1998b; Reid and Thoms, 2008) and although they do not map directly onto one another (Clifford *et al.*, 2006; Harvey and Clifford, 2008), functional habitats are associated with assemblages of flow types (Harvey *et al.* 2008a). Rowntree (1996) developed a biotope matrix in collaboration with stream ecologists to classify physical habitats based on substrate (a component of functional habitats) and flow type (a component of biotopes) (**Figure 2.4**).

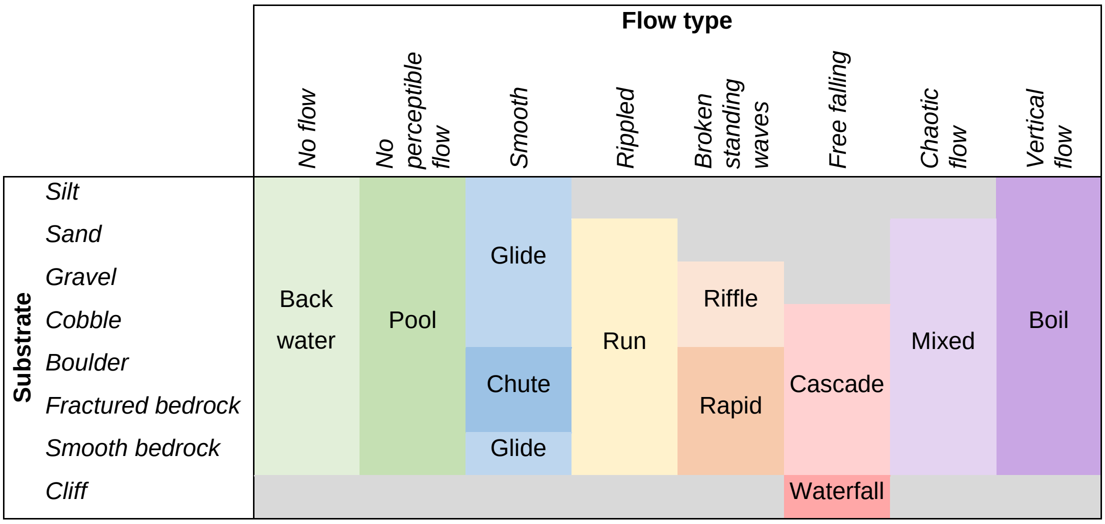


Figure 2.4. *The biotope matrix describing physical habitats as a function of substrate and flow type classes. Reprinted Figure 4.1 from* Rowntree (1996, p.47).

As well as individual habitat types being important for stream ecology, a diverse range of physical habitats may be more important in some ecological communities (Gilvear and Willby, 2006). For example, habitat diversity provides a range of habitats for different life stages of individual species, for mobile species, and a range of smaller organisms confined to a single habitat as prey for larger species (Woodward and Hildrew, 2002). Physical habitat diversity is also used as an indication of conservation value from a management perspective (Raven *et al.*, 1996), can be used to assess restoration success (Milner and Gilvear, 2012) and the impacts of different engineering approaches (Gostner *et al.*, 2013). Therefore, this thesis considers components of both functional habitats and biotopes to explore the response of physical habitat type and diversity to catchment-level effects.

2.1.3 Which catchment characteristics impact physical habitats in rivers?: An Evidence Review

This thesis cannot capture every element of the catchment system outlined in **Figures 2.1** and **2.2** in detail, but it can focus on areas that are particularly poorly understood. To identify these gaps in understanding, an evidence review is conducted utilising a quick scoping methodology with the aim of gaining an objective assessment of which catchment characteristics are frequently used to explain patterns of physical habitats. The review focuses on large-scale data-science studies investigating multiple reaches.

2.1.3.1 Quick scoping review methodology

To explore the literature, a quick scoping review is conducted following the protocol laid out by the Joint Water Evidence Group (Collins *et al.*, 2014). A quick scoping review provides an assessment of the *"size and type of evidence available"* (Collins *et al.*, 2014, p. 2) and does not offer a critical appraisal of the evidence, appropriate for the task of identifying research gaps.

The quick scoping methodology requires the development of a protocol with a number of steps: (i) identification of a key question (and sub-questions) and the scope of the review (e.g. language of papers, time period etc.); (ii) focusing the question using PICO elements (Population, Intervention, Control and Outcome); (iii) evidence search using keywords identified from the PICO elements; (iv) screening the search results to narrow down the literature using inclusion and exclusion criteria; and (v) information extraction systematically from the selected papers to create an evidence database (Collins *et al.*, 2014).

(i) Key question

Following this methodology, the key question is *"Which catchment characteristics affect physical habitats in rivers?"*. The specific sub-questions are: (i) which catchment characteristics are frequently being employed to assess the catchment-level effect on physical habitats in studies assessing multiple sites? and (ii) how effective are catchment-level characteristics in explaining physical habitat features in comparison to reach-level characteristics? Only papers in English are selected with no regional or time period restrictions.

Table 2.1. PICO elements and exclusion criteria used for the quick scoping review

PICO	Inclusion	Exclusion	Other exclusion criteria
Population	• Rivers/streams • Catchments	• Reservoirs • Lakes	• Review papers • Papers assessing effectiveness of management technique • Methods papers • Decision support systems
Intervention	• Catchment level characteristics	• Not catchment level characteristics • Only single characteristics • Future climate scenarios	
Control	• Reach level characteristics	• Only reach level characteristics	
Outcome	*Physical habitat features* • Reach morphology • Substrate • Macrophytes • Large wood *Or habitat indicators* • Invertebrates	• Runoff • Flow regimes • Pollutants/pesticides • Genetics • Water quality/temperature • Modelled response variables • Bank erosion	

(ii) PICO elements

The PICO elements used to define the question and build the search criteria are in **Table 2.1**. The question focuses on rivers and their catchments and is specifically interested in catchment-level characteristics. Therefore, characteristics not at the catchment-level (e.g. discharge at a site or land cover within a buffer strip) are not considered. However, many studies also consider reach-level characteristics and therefore studies that include additional characteristics at smaller scales are retained as a control measure. The outcome, or response variable, of the studies may not be termed 'physical habitat' in the papers so multiple features are included as well as macroinvertebrates which are often related to physical habitat (see Section 2.1.2 for references). Papers concerning fish and other mobile biota are excluded as they travel throughout the system so cannot be directly related to physical habitats within a specific reach. Review papers and modelling studies are excluded so only empirical studies remain.

(iii) Evidence search

Boolean operators are used to search for papers (**Box 1**). Many of the exclusion terms (indicated by NOT) were identified through trial and error. The search was conducted initially on 5th January 2018 and updated on 6th June 2019 using Web of Science search engine producing 480 potential papers.

Box 1: Boolean operators are used to streamline the search results
"catchment characteristic" OR "catchment scale" OR "catchment-scale"*
AND river OR stream**
AND habitat OR ecolog* OR *morph**
NOT groundwater
*NOT flood**
*NOT landslid**
*NOT reservoir**
*NOT lake**

(iv) Screening

Papers were screened by assessing the relevance of the paper to the PICO requirements (**Table 2.1**) by reading first the title, then the abstract and finally the full paper. In total 29 papers were selected, with 47 rejected for focusing solely on land cover rather than multiple catchment characteristics. The 29 papers selected are listed in **Appendix 2A**.

(v) Information extraction

Most papers considered physical habitat as a primary response variable, however, five papers considered physical habitat in addition to outcomes such as water quality and fish. The results discussed below relate only to findings of the papers relating to the outcome of interest, physical habitats, macroinvertebrates and other sedentary benthic organisms (e.g. diatoms and bivalves). The significance of the effect of an individual characteristic was determined if the paper included the characteristic in their final statistical model or if they reported significance test results for individual characteristics. Whether catchment or reach-level characteristics were deemed more important was based on the conclusions drawn by each paper. Not all papers considered catchment and reach characteristics individually so no judgement could be made in these cases.

2.1.3.2 Key patterns in the literature identified from evidence review

Catchment-level characteristics such as topography (e.g. mean elevation, slope and shape of the catchment), network (e.g. drainage density), climate, soil, geology, land cover and other characteristics (e.g. the density of lakes or roads) have all been related to physical habitat response variables (**Figure 2.5**). Many papers also include measures of the reach's location within the catchment, such as catchment area upstream of a reach, the spatial coordinates of the reach, stream order, distance along the network and elevation of the reach (e.g. Mugodo *et al.*, 2006; Parsons and Thoms, 2007; Frappier and Eckert, 2007). The position of the reach in relation to the longitudinal gradient of the catchment is deemed to be an important control on channel character (see Section 2.1.1.1 for literature and **Figure 2.2**). Most papers also include reach-level characteristics such as local slope (e.g. Davies *et al.*, 2000; Thompson *et al.*, 2008; Lindholm *et al.*, 2018), sinuosity (e.g. Townsend *et al.*, 2003; McRae *et al.*, 2004; Jähnig *et al.*, 2015) and the presence of large wood (e.g. King *et al.*, 2012; de Castro *et al.*, 2017).

Land cover is the most frequently considered catchment characteristic (**Figure 2.5**) and was solely considered by many papers excluded from this review. This indicates a trend in research focusing on anthropogenic impacts on river reaches. This is because urban and arable land covers are related to a range of reach-level modifications such as fine sedimentation, changes in flow, dredging and channel straightening (Brookes, 1988; Sear *et al.*, 2003; Walsh *et al.*, 2005). The management of such anthropogenic pressures has therefore become a priority for catchment management practices so is a popular research topic. In 74% of papers that considered it, land cover had a significant impact on the response variable making it an influential control on physical habitat. For example in Manfrin *et al.*'s (2016) study, the proportion of urban land and woodland in the catchment were the most important factors in all ten models explaining the biotic index in 95 different streams in Italy, compared to the mean elevation and slope of the catchment which were only important in three and two of the models respectively.

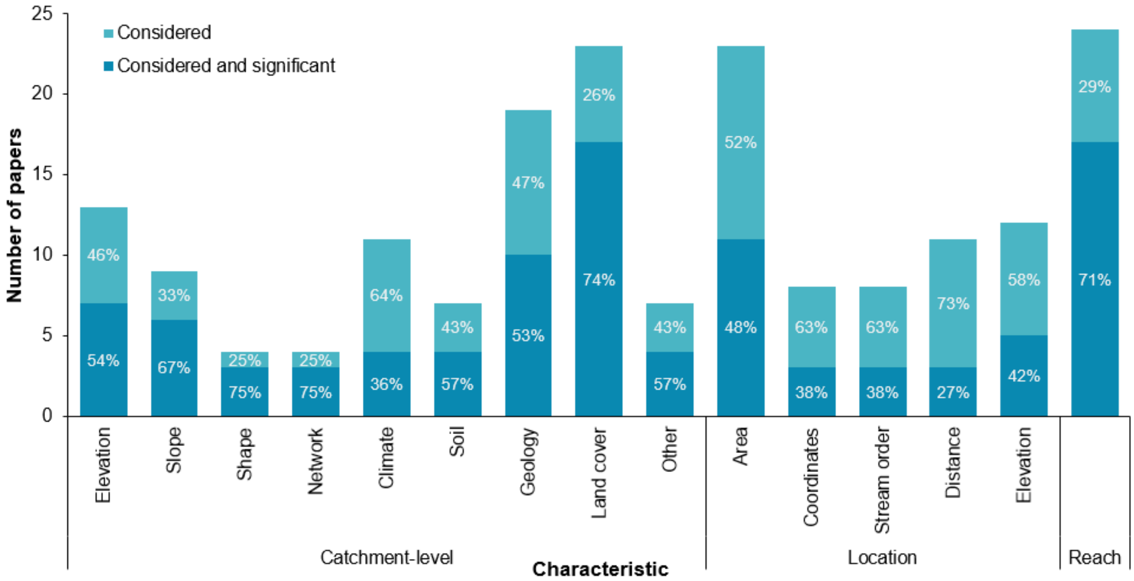


Figure 2.5. *Number of papers that consider each catchment characteristic. Bars are split according to which percentage these papers showed significant influence on the response variable for each characteristic.*

Other catchment characteristics – such as geology, elevation, soil and climate – are included frequently in studies but only significantly influence physical habitats in around half of papers. Climate characteristics (most frequently precipitation) are not often significant (e.g. Death and Joy, 2004; Mugodo *et al.*, 2006; Benone *et al.*, 2017), despite their inclusion in over a third of papers assessed in this review. Similarly, most papers include some measure of the location of the reach in the catchment (>85% of papers), most frequently area, but location characteristics are frequently not found to be significant for physical habitat. This is despite the supposed importance of the position of the reach along the longitudinal gradient of the catchment (**Figure 2.2**).

Morphometric properties of the catchment and network are less frequently considered by papers (**Figure 2.5**), with only four papers considering measures of catchment shape (such as elongation ratio and basin length; Davies *et al.*, 2000; Townsend *et al.*, 2003; Woodcock *et al.*, 2006; Parsons and Thoms, 2007) and network structure (such as bifurcation ratio, total stream length and drainage density; Davies *et al.*, 2000; Townsend *et al.*, 2003; Feld, 2004; Hutchens *et al.*, 2009). However, shape and network characteristics are both found to have significant influence on physical habitat in 3 out of 4 papers (**Figure 2.5**). Perhaps this is because these characteristics reflect the hillslope and network morphology of the catchment which is related to flow and sediment regimes (Gregory and Walling, 1973; **Figure 2.1**).

Reach characteristics were also frequently considered by most papers (83%) and frequently found to have a significant influence on physical habitat features (**Figure 2.5**). However, most papers showed that catchment-level variables had a greater impact on the physical habitat than reach-level variables, with only 10% of papers reporting that reach-level variables had a greater impact (**Figure 2.6**). Many papers also reported that both catchment and reach-level variables are important for the response variable which is understandable due to the nested hierarchical nature of the river system meaning that properties at the regional, catchment, segment and reach levels influence meso-level physical habitats (**Figure 2.1**). Interestingly, despite many of the studies acknowledging the importance of scale in their manuscripts, few use statistical methods that account for the hierarchical nature of the characteristics on the habitat response with most using linear regression models or Principal Components Analysis (PCA).

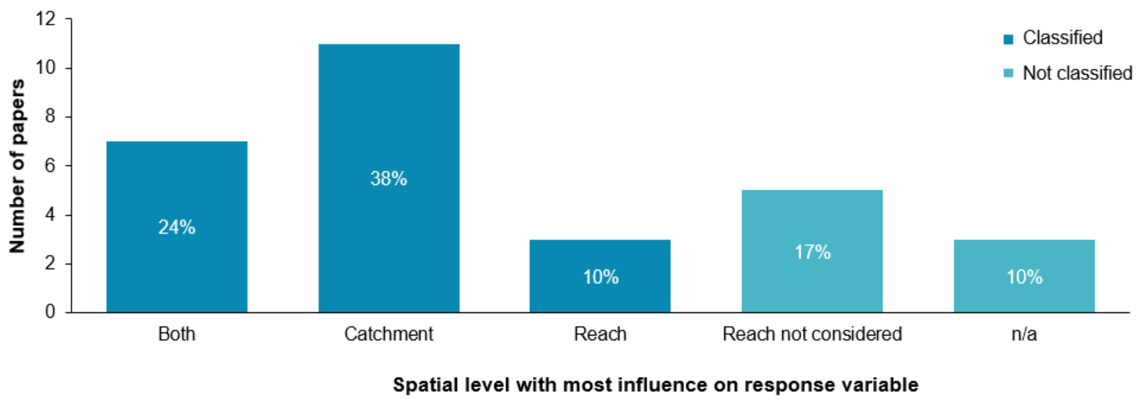


Figure 2.6. *Spatial level of characteristics that have most impact on physical habitat response variables in classified papers. n/a indicates papers where no judgement on whether the reach or catchment-level had more influence on the physical habitats could be determined.*

In summary, catchment characteristics are shown by this review to be more, or as, important as reach characteristics in influencing physical habitats (**Figure 2.6**). Land cover is the focus of most studies due to the anthropogenic pressures associated with it (Sear *et al.*, 2003; Walsh *et al.*, 2005). However, the main finding of the quick scoping review is a disparity between the characteristics that are assumed to be important (e.g. downstream location, elevation, climate and geology) and those that are frequently shown to have a significant effect (e.g. catchment and network morphometry). The neglect of catchment and network morphometry excludes key hydrological processes from catchment studies that connect regional and catchment-level characteristics to sub-segment level systems (**Figure 2.1**). It is therefore suggested that studies trying to understand catchment-level controls on physical habitats are frequently focusing on anthropogenic pressures rather than holistically accounting for catchment-level effects. Therefore, this thesis considers a range of catchment characteristics, including both frequently used (e.g. land cover) and overlooked (e.g. network structure) characteristics to better represent holistic catchment-level effects.

2.2. DATA-SCIENCE FOR BROAD-SCALE RESEARCH

Data-science is a blend of statistics and computer science developed to confront the increasing amount of 'big-data' available and the move towards open source, reproducible methods in scientific research (Blei and Smyth, 2017). Therefore, data-science methods have been embraced across the sciences; including ecology, hydrology and geomorphology to understand complex landscape interactions (Farley *et al.*, 2018; Peters *et al.*, 2018). Data-science methods can be met with reluctance as the focus of research moves towards building increasingly accurate and complex models using 'black-box' methods, with less focus on scientific theory (Karpatne *et al.*, 2017). However, theory-guided data-science is possible when models consider scientific theory and care is taken to explain variable interactions rather than ever seeking to optimise model accuracy (Karpatne *et al.*, 2017; Mac Nally, 2000).

A data-science methodology is adopted by this thesis, utilising both traditional statistics and machine learning approaches to identify catchment-level effects on river habitats. It is used, with an understanding of scientific theory, for broad-scale research following a practice of using monitoring datasets for scientific enquiry. This allows for research to be conducted over wider spatial and temporal scales rather than focusing on individual case studies. The benefits and limitations of a broad-scale approach are discussed below followed by a description and critique of the monitoring dataset used in this thesis, the River Habitat Survey (RHS).

2.2.1 Utilising broad-scale monitoring data

Broad-scale datasets, extensive monitoring datasets collected by regulatory bodies for policy compliance, offer the opportunity to conduct research at wider spatial and temporal scales that are necessary to understand the variations in river functioning over scales that are relevant for effective management. To collect the data necessary to conduct such broad research first-hand would require significant time and resources often unavailable to researchers.

There are vast amounts of empirical monitoring data available that is under explored but which could be adapted to answer scientific questions such as linking habitat responses to physical controls. **Figure 2.7** shows the range of hydrological data and hydromorphology assessments across the globe, highlighting the dominance of data availability in Europe and the US.

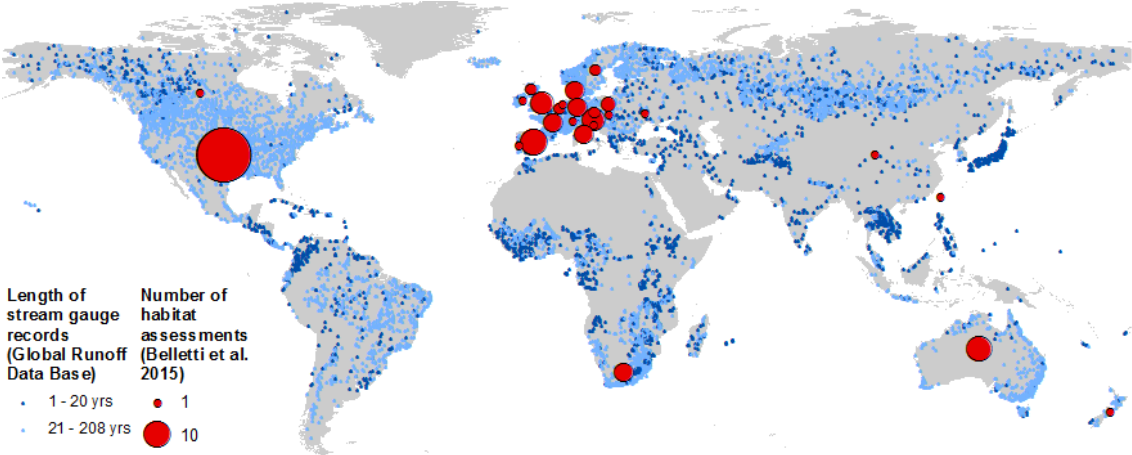


Figure 2.7. *Examples of broad-scale datasets at a global level including the number of hydromorphological assessment methods available in each country (Belletti et al., 2015) and the location of stream gauges (GRDC, 2017) with the length of discharge data available at each gauge.*

With 121 hydromorphology survey methods identified from countries across the world by Belletti *et al.* (2015) (**Figure 2.7**), there is the potential for numerous physical habitat datasets to be available to researchers. There is also extensive ecological data available with many examples of national biomonitoring programmes employed in Europe, USA, China, Korea and other countries, measuring a range of ecosystem indicators including macroinvertebrates, fish, aquatic plants and benthic diatoms (Park and Hwang, 2016). These measurements are often used to derive a judgement on the ecological quality of a waterbody as part of a legislative requirement. However, the original monitoring data could be used to explore ecohydrological interactions over larger spatial and temporal scales. Other datasets such as the NRRI and EU Wiki are available for exploring regional patterns of restoration measures using broad-scale analysis (Smith *et al.*, 2014).

There are challenges in using existing datasets, that introduce biases into the data. The purpose of the monitoring is a key bias, determining the variables recorded and site selection which may be driven by alternate motives rather than a rigorous statistical approach (Wessels *et al.*, 1998). This is important because the content of the dataset defines the scope of the study, limiting the questions that can be answered using broad-scale analysis. For example, broad-scale data may only be available in certain regions of the world, and this combined with data accessibility may limit the ability of broad-scale data to be used to explore certain biomes or time periods (**Figure 2.7**). However, the development of data-science methods and GIS expand the questions that can be addressed using broad-scale data to achieve interesting and low-cost results. Data-science and GIS are used in this thesis to capture catchment-level effects on physical habitats, monitored by the national River Habitat Survey.

2.2.2 The River Habitat Survey

The River Habitat Survey (RHS; Raven *et al.* 1996) dataset is used here to identify physical habitat features. The RHS was developed in 1994 and records over 100 habitat features at each site with over 24,000 sites sampled across England and Wales to date (Naura *et al.*, 2016).

At each site, data are collected within a 500m reach involving ten 'spot-checks' along cross-channel transects at 50m intervals recording channel and bank features. A 'sweep-up' is also conducted to record features and modifications that occur in over a third of the 500m reach. Data are collected by certified practitioners after attending a training course, and audit checks suggest there are few non-sampling errors in these data (Fox *et al.*, 1998).

The RHS has frequently been used to conduct broad-scale research, most commonly to identify relationships between individual habitat features and map-derived data, particularly location within the catchment (Harvey and Wallerstein, 2009; Jeffers, 1998a; Naura *et al.*, 2016; Vaughan *et al.*, 2013), to predict ecological populations from habitat features (Naura and Robinson, 1998; Vaughan *et al.*, 2007) or to create river typologies (Harvey *et al.*, 2008a; Jeffers, 1998a; Jusik *et al.*, 2015). Broad-scale datasets lend themselves to answering such questions and tasks as the datasets contain surveys across a range of environmental conditions.

2.2.2.1 Methodological considerations of the RHS

The RHS was developed to assess river quality by regulatory bodies so has biases and limitations associated with it. RHS sites were selected via a random stratified sampling approach, with three sites randomly selected along the network within each Ordnance Survey (OS) 10 x 10 km grid square of the UK (Jeffers, 1998b). During the 1995-1997 baseline survey, the network was based on the OS 1:250,000 map which excluded headwater streams but post-2003 this was corrected by randomly assigning two sites to headwater streams on the OS 1:50,000 map. Sites are not revisited so change cannot be detected at individual sites. However, the sampling strategy and high volume of data allow similar sites to be compared (Environment Agency, 2010). The site selection methodology also means that some catchments do not have multiple sites on the same stream, making upstream downstream comparisons difficult within individual rivers (Seager *et al.* 2012). RHS is also not designed for large, deep rivers so physical habitats in these environments are not measured. This means the RHS is not truly representative of the entire catchment.

The majority of RHS critiques are because of its quick, bankside survey design. This means that detailed measurements of process are not collected but inferred from observations of habitat features. Also, methodological procedures such as selecting the dominant feature at each transect means that the diversity of the features are not captured within spot-checks.

Surveyor error or biases also impact the survey, for example, some features such as sand and bedrock are hard to identify from the bankside. The surveys are designed to be conducted on smaller streams as a 500m reach captures different ranges of habitat units, for example, 500m in a small river may capture recurring pool-riffle sequences compared to a large river that may exhibit only part of a pool-riffle sequence within the reach length (Emery *et al.*, 2004). The surveys are also designed to be conducted in spring at low flow, so substrate is still visible and macrophytes identifiable. However, these recommendations are not always observed by surveyors.

Whilst there are limitations to the data, the volume of information recorded means studies identifying general relationships of river habitats and morphology can be conducted at the large-scale (e.g. Newson *et al.* 1998; Emery *et al.* 2004; Harvey *et al.* 2008b). There are also extensions to the RHS such as the Urban River Survey which observes more features relevant to urban systems (Davenport *et al.*, 2004). Another survey, MoRPh, adapts the RHS methodology to minimise some of its key limitations by varying survey reach length based on channel width, and recording all features observed to capture habitat diversity, along with additional variables of interest (Shuker *et al.*, 2017). Whilst MoRPh surveys would be useful here, at the time of inception of this thesis the method was in its infancy and datasets were not extensive enough to conduct such broad-scale research. However, the number of surveys has grown quickly to over 3,000 MoRPh surveys (as of March 2020) due to the uptake of MoRPh surveys by citizen scientists (Gurnell *et al.*, 2019), and the inclusion of MoRPh in the Biodiversity Metric 2.0 (Crosher *et al.*, 2019) as a tool to assess the distinctiveness and condition of habitats.

Table 2.2. *Physical habitat indices calculation: (a) Features observed in the RHS used for index calculation; (b) Indices extracted from the RHS observations in this thesis to reflect physical habitat diversity and average habitat type within each 500m reach; (c) Regulatory scores calculated externally.*

(a)

Feature	Description	Abbreviation
Flow types		
Free-fall	Count of dominant free fall spot checks	FF
Chute	Count of dominant chute spot checks	CH
Chaotic flow	Count of dominant chaotic flow spot checks	CF
Broken wave	Count of dominant broken wave spot checks	BW
Unbroken wave	Count of dominant unbroken wave spot checks	UW
Ripple	Count of dominant ripple spot checks	RP
Smooth	Count of dominant smooth spot checks	SM
Upwelling	Count of dominant upwelling spot checks	UP
No perceptible	Count of dominant no perceptible flow spot checks	NP
Dry	Count of dominant dry spot checks	DR
Sediment types		
Bedrock/boulder	Count of dominant bedrock and boulder spot checks	BO
Cobble	Count of dominant cobble spot checks	CO
Gravel-pebble	Count of dominant gravel-pebble spot checks	GP
Sand	Count of dominant sand spot checks	SA
Silt	Count of dominant silt spot checks	SI
Clay	Count of dominant clay spot checks	CL

(b)

Index	Calculation	Units
Flow type diversity	$= \sum \left(\frac{n}{N}\right)^2$ n= count of each flow type. N = total number of flow types at the site.	0-1 (1=most diverse)
Sediment diversity	$= \sum \left(\frac{n}{N}\right)^2$ n= count of each sediment type. N = total number of sediment types at the site.	0-1 (1=most diverse)
*Average reach sediment size**	$= \frac{(-8 \cdot BO - 7 \cdot CO - 3.5 \cdot GP - 1.5 \cdot SA + 1.5 \cdot SI + 9 \cdot CL)}{(BO + CO + GP + SA + SI + CL)}$ Sediment type abbreviations represent the number of spot checks allocated to each sediment size class	Approx. Phi scale (9 = coarse) *
Average reach flow type speed	$= \frac{(0 \cdot DR + 1 \cdot NP + 2 \cdot UP + 3 \cdot SM + 4 \cdot RP + 5 \cdot UW + 6 \cdot BW + 7 \cdot CF + 8 \cdot CH + 9 \cdot FF)}{(DR + NP + UP + SM + RP + UW + BW + CF + CH + FF)}$ Flow type abbreviations represent the number of spot checks allocated to each flow speed class	Flow type speed scale (9 = fast)

**Sediment size index inverted so the highest values are coarsest for ease of interpretation. However, the indices are not inverted in the paper in Chapter 4 to be consistent with the commonly used Phi scale.*

(c)

Index	Calculation	Units
Habitat Quality Assessment (HQA)	Sum of assigned scores for flow types, channel and bank features, bank vegetation, tree cover, and land use.	HQA scale (high = good quality)
Habitat Modification Score (HMS)	Sum of assigned scores for modifications to the channel bed and banks and the presence of structures such as culverts, bridges, weirs and dams.	HMS scale (high = modified)

2.2.2.2 *Extracting habitat indices from the RHS dataset*

Before physical habitat indices can be extracted from the RHS, the dataset is quality controlled. RHS survey protocol has changed since its initial implementation in 1994. This causes slight changes in the information collected but has little effect on the variables of interest in this thesis, causing only the names of the flow types to change so they are referred to with 2003 terminology (Environment Agency, 2003).

There is also missing data. Of the 21,886 RHS surveys in England from 1994 to 2016 at the start of this research, 19,258 surveys collected dominant flow and sediment type. Both flow type and sediment observations should have ten observations per survey. Where observations did not equal ten, the variable at that site was removed from the analysis to prevent error.

Each RHS site is assigned a National Grid Reference so it can be input to the GIS as point data. Sites were snapped the nearest link on the 1:50,000 blueline network (Moore *et al.*, 1994), the network used to sample the sites, within 500m using the RivEX 'Snap' add-on tool to ArcGIS (Hornby, 2010). Sites over 500m from the network were removed so sites are not linked to the incorrect river.

Flow type and sediment are the focus of this thesis to reflect biotope and functional physical habitat types respectively (see **Figure 2.4** and Section 2.1.2 for details). The dominant flow type and sediment class is recorded for each of the ten spot-check cross sections (**Table 2.2a**).

Table 2.3. *Interpretation of the average reach habitat indices: (a) flow type speed; (b) sediment size according to Phi scale and; (c) sediment size on inverted Phi scale.*

(a)

Flow type speed	**Index value**
Free-fall	9
Chute	8
Chaotic	7
Broken waves	6
Un-broken wave	5
Rippled	4
Smooth	3
Upwelling	2
No perceptible	1
Dry	0

(b)

Sediment size (Phi scale)	**Index value**
Boulder	-8
Cobble	-6
Pebble	-2
Coarse sand	0
Fine sand	3
Coarse silt	5
Fine silt	7
Clay	>8

(c)

Sediment size (inverted Phi scale)	**Index value**
Boulder	8
Cobble	6
Pebble	2
Coarse sand	0
Fine sand	-3
Coarse silt	-5
Fine silt	-7
Clay	<-8

The observations are transformed into indices of physical habitat representing the average type and diversity of dominant habitats for each 500m survey reach. Habitat type indices are based on indices derived in the Urban River Survey (Davenport *et al.*, 2004) where each sediment class is assigned number representative of grainsize (according to the Phi scale), and each flow type is assigned an incremental number reflecting an approximate flow velocity gradient (**Table 2.2b**).

Sediment size and flow type speed values are averaged for the reach, although the sediment size index is inverted to give the coarsest sediments the highest values to simplify interpretation. Flow type speed index does not represent true velocity as flow types reflect both velocity and depth (Zavadil *et al.*, 2012). Therefore, flow types such as free-fall and chute are given higher velocities even though larger rivers may be faster flowing. Diversity of flow type and sediment classes was calculated using Simpson's diversity index (Simpson, 1949) (**Table 2.2b**). The average habitat indices can be interpreted according to **Table 2.3**.

Two scores are calculated automatically in the RHS database: Habitat Quality Assessment (HQA) and Habitat Modification Score (HMS) (**Table 2.2c**). HQA reflects the quality and diversity of the river corridor whereas HMS reflects the extent of modification in the channel. Features observed by the RHS survey that are relevant to each measure are assigned a score which is weighted by expert opinion and summed to calculate the score (for more details see Raven *et al.*, 1998). These scores are included as they reflect multiple features of the reach, rather than solely physical habitat attributes. They are also used for river quality assessment by practioners so it will be useful to see how these measures respond to catchment-level effects.

The majority of previous studies using the RHS dataset have conducted analysis using subsets of the data, using only the most natural sites (Harvey *et al.*, 2008a, 2008b; Naura *et al.*, 2016) or for certain time periods (e.g. 1994-1996, Jeffers, 1998a; 2007-2008, Vaughan *et al.*, 2013). However, this work uses the whole dataset, post-quality control, to explore catchment-level effects at all sites, including modified sites, over a longer time frame.

2.3 SUMMARY AND SPECIFIC RESEARCH OBJECTIVES

This thesis investigates how catchment-level effects influence physical habitats at the reach-level. Specific research objectives are defined to assess this broad goal and are outlined in **Table 2.4** with reference to the specific aim (Chapter 1) they address and the chapter in which they are investigated.

Based on the material presented in this chapter, it is clear that a range of regional controls influence the discharge and sediment that moves through the spatial hierarchy to influence river reaches (**Figure 2.1**; Section 2.1.1.1). The evidence review highlights that most previous large-scale studies concentrate on the land cover catchment characteristic (Section 2.1.3). This focus on anthropogenic pressures is also reflected by integrated catchment management plans which fail to consider holistic catchment-level effects (Section 2.1.1.2). Therefore, this thesis considers a range of catchment characteristics for creating a typology of catchment-level effects across England and Wales, to explore interactions between both natural and anthropogenic effects and their spatial distribution (see objectives for Chapter 3; **Table 2.4**).

Table 2.4. *Aims addressed by each objective (chapter number and objective letter in brackets).*

Aim 1	Aim 2	Aim 3	Objective
			(**3A**) To build a typology of catchment-level effects that is practically useful for implementation by river managers.
			(**3B**) To explore how effective a typology of catchment-level effects is at explaining physical habitats in river reaches.
			(**4A**) To quantify network topology within catchments by creating a metric fit for multiple disciplinary use.
			(**5A**) To identify how important confluences influence the effect of network topology on river habitats.
			(**5B**) To investigate which properties of upstream tributaries influence confluence importance.
			(**5C**) To explore how catchment morphometry influences the effect of network topology on river habitats.
			(**6A**) To use a range of GIS-derived catchment and river properties at different hierarchical levels, to explain patterns of physical river habitats.

Aim 1: *Integrate a theoretical understanding of the river network with regulatory monitoring data to identify catchment-level effects on river habitat.*
Aim 2: *To improve understanding of the roles network and catchment structure can play in assessing habitat distribution within and between catchments.*
Aim 3: *Apply knowledge of catchment-level effects to explain national patterns of reach-level habitats and explore management applications.*

The evidence review in Section 2.1.3 identifies a key gap in previous studies; the under representation of catchment and network morphometry effects. Theoretically, these are critical dimensions of the catchment that route catchment-level effects downstream through the network to influence broad habitat patterns (**Figure 2.2**; Section 2.1.1.1). Therefore, this thesis focuses on network structure, exploring the effects of the entire network and individual confluences on habitats (see objectives for Chapters 4 to 5; **Table 2.4**).

As the spatial hierarchy indicates, wider catchment processes influence the meso-level (**Figure 2.1**), the level at which physical habitats are conceptualised. This thesis focuses on the flow type and sediment features of physical habitats as they relate directly to concepts of biotope and functional approaches (**Figures 2.3 and 2.4**) that have been shown to reflect the biotic and abiotic conditions of the channel. Not only the type of habitat but also the diversity of habitats available in a reach (Section 2.1.2) are considered and extracted from the RHS (**Table 2.2**). The use of the RHS, or any broad-scale dataset, has limitations (outlined in general in Section 2.2.1 and specifically related to the RHS in Section 2.2.2.1), but these are deemed acceptable in relation to the volume of data points that allow analysis to be conducted at the national-level.

Associations between habitats and catchment-level effects will have application to river management. Management interventions such as reach restoration, are often conducted opportunistically rather than strategically without consideration for catchment-level effects which may be detrimental to the success of individual projects and broader improvements in river health (Section 2.1.1.2). The objectives of this thesis consider the application of this work to aid the rapid and holistic assessment of catchment-level effects for river managers. For example, the catchment-level effects typology is designed to be a practical tool (Objective 3a), the metrics intended to quantify network structure are simple and easy to extract for any catchment (Objective 4a). In addition, all analysis is conducted using a dataset collected by regulators that gathers data of importance to managers using transferable methodologies (**Table 2.4**).

The objectives in **Table 2.4** are repeated in the introduction of each chapter and results pertaining to each specific objective are assessed in the conclusion of the relevant chapter. The conclusions relating to the overall aims of the thesis are discussed in Chapter 7.

CHAPTER 3

A waterbody typology derived from catchment controls using self-organising maps

3.1 CHAPTER INTRODUCTION

This chapter was published in the journal Water in December 2019 and is included as the submitted manuscript in Section 3.2. The gaps in knowledge and thesis objectives addressed by the paper are briefly summarised below. For more detailed explanation, see Section 3.2.

Previous catchment studies (see evidence review in Section 2.1.3) and integrated catchment management often focus on anthropogenic controls on river reaches. This is problematic as they do not account for other aspects of the catchment, such as geology, climate and morphometry, which influence catchment function and its effect on river reaches. It is challenging to analyse multiple and complex catchment characteristics as they often co-correlate and interact. This chapter uses a machine learning data-science approach, self-organising maps (SOMs), to account for the complexities of multivariate data and creates a typology to simplify the functionality of the catchment system. Typology creation is commonplace in river science, but there are few typologies that consider catchment-level effects, and fewer that have created a typology of solely GIS-derived catchment-level effects. The objectives of the chapter are:

Objective 3a: To build a typology of catchment-level effects that is practically useful for implementation by river managers.

Objective 3b: To explore how effective a typology of catchment-level effects is at explaining physical habitats in river reaches.

The GIS-derived nature of the typology means that it can be applied continuously across England and Wales making it a useful tool for rapid assessment of catchment-level effects. The typology is also evaluated against broad-scale River Habitat Survey dataset indices (**Table 2.2**) to demonstrate the usefulness of the typology for reach-level studies. Work not included in the paper is presented in Section 3.3 including (i) a comparison of the machine learning approach used in the paper to another statistical approach; and (ii) a potential application of the typology for catchment complexity assessment. Additional methodological justification of the number of clusters and the selected SOM grid size is in **Appendix 3A** and **3B**. The code to produce the typology is in **Appendix 3C**.

3.2 PUBLISHED PAPER

Text, tables and figures copied directly from Heasley et al. *(2020). Numbering of sections and figures changed to coincide with the thesis. The appendix to the paper is* ***Appendix 3A*** *in the thesis. Citations in the published paper are numbered to agree with the journal's formatting but are included as full citations in this thesis. The reference list for the paper is included in the full reference list for the thesis. The published paper is accessible online at* *https://doi.org/10.3390/w12010078*

3.2.0 Abstract

Multiple catchment controls contribute to the geomorphic functioning of river systems at the reach-level, yet only a limited number are usually considered by river scientists and managers. This study uses multiple morphometric, geological, climatic and anthropogenic catchment characteristics to produce a single national typology of catchment controls in England and Wales. Self-organising maps, a machine learning technique, are used to reduce the complexity of the GIS-derived characteristics to classify 4,485 Water Framework Directive waterbodies into seven types. The waterbody typology is mapped across England and Wales, primarily reflecting an upland to lowland gradient in catchment controls and secondarily reflecting the heterogeneity of the catchment landscape. The seven waterbody types are evaluated using reach-level physical habitat indices (including measures of sediment size, flow, channel modification and diversity) extracted from River Habitat Survey data. Significant differences are found between each of the waterbody types for most habitat indices suggesting that the GIS-derived typology has functional application for reach-level habitats. This waterbody typology derived from catchment controls is a valuable tool for understanding catchment influences on physical habitats. It should prove useful for rapid assessment of catchment controls for river management, especially where regulatory compliance is based on reach-level monitoring.

3.2.1 Introduction

Geomorphic functioning of rivers is nested within a hierarchy of levels, each with progressively broader extents from sub-reach ($<10^1$ m), reach ($\sim10^1$ - 10^2 m), segment ($\sim10^2$ - 10^3 m) to catchment levels ($>10^3$ m) (Frissell *et al.*, 1986). River managers often focus on individual reaches, yet functioning is ultimately controlled by the boundary conditions of the catchment (Brierley and Fryirs, 2000; Kondolf *et al.*, 2003) so that 'in every aspect the valley rules the stream' (Hynes, 1975, p.12). This paper develops a typology of catchment controls that influence river reaches, within sub-units of catchments referred to as waterbodies.

The hierarchical explanatory framework approach described by Frissell *et al.* (1986) and others (Gurnell *et al.*, 2016) has been adopted by river scientists and mangers. This has led to the widespread acceptance that knowledge of multidisciplinary, multiparameter controls that influence process must be incorporated within catchment management (Beechie *et al.*, 2010; Church and Ferguson, 2015; Downs and Gregory, 2004; England and Gurnell, 2016). However, multiple controls are not frequently fully integrated within management because gradients of anthropogenic land use are often superimposed onto the underlying properties of the natural landscape, making natural features of the catchment that influence river function more difficult to identify (Allan, 2004). Multiple catchment controls are considered by some previous river typologies designed for river management, for example, using catchment controls such as geomorphology, geology, climate, and land cover for river section delineation (e.g. River Styles typology for Australia, Brierley and Fryirs, 2000; REFORM typology for Europe, Rinaldi *et al.*, 2016). However, these typologies use individual catchment controls in isolation to define homogeneous reaches rather than capturing associations between controls to explore their spatial distribution. How multiple catchment controls may best be incorporated into typologies should be explored to allow for improved integrated catchment management.

We aim to produce a waterbody typology derived from catchment controls, that combines multiple catchment characteristics into a practical set of types that are scientifically robust and useful for management decision-making. Defined by the Water Framework Directive (WFD), waterbodies are sub-units of catchments designed to contain rivers of similar condition and are used to assess WFD ecological and chemical quality targets according to European standards (European Commission, 2000). Waterbodies are a commonly applied delineation of the landscape as they are meaningful to river management (Acreman *et al.*, 2008). The waterbody typology developed here should capture a wider range of catchment controls that influence reach-level features than is usually considered by catchment management or existing river typologies. The presence of numerous and complex catchment controls presents a challenge for analysis and interpretation, so a machine learning technique, self-organising maps (SOMs), is employed to derive the typology from the large multivariate dataset. The typology captures the dominant catchment controls that influence river reaches across numerous waterbodies in England and Wales, rather than directly classifying reach processes and features. The patterns identified from a typology that represents controls on reach-level features should aid broad-level and strategic management (as opposed to management at an operational level), by encouraging wider appreciation of multiple catchment influences on river reaches.

3.2.1.1 Approaches to typology creation in river research

Characterisation of river types is a frequent occurrence within river studies, with over 100 river typologies developed over the past 125 years (Naura *et al.*, 2016). Both scientific and management driven approaches for typology development have the same fundamental aim: to reduce the complexity of the river system to a practically useful set of types (Kondolf *et al.*, 2003). Yet their use differs; scientific approaches use typologies to explore the distribution of homogeneous classes and identify natural thresholds whereas applied approaches use typologies to identify reference sites and to improve communications between disciplines and stakeholders using simple classifications (Kondolf *et al.*, 2003; Tadaki *et al.*, 2014).

Classifications are often critiqued for not accounting for enough variation, being over-simplified and drawing arbitrary boundaries on natural continuums (Wright *et al.*, 1984). Issues also arise when a classification becomes a guiding principle and our understanding of a river becomes limited to a 'type' when additional factors will also impact the management approach appropriate for a reach (Kondolf *et al.*, 2003). However, by recognising a typology as a tool that is 'an abstraction of what would otherwise be an inconceivable array of natural variation' (Tadaki *et al.*, 2014, p.362) and by not pushing it beyond its design, these limitations may be accounted for.

River classification may be achieved by either a *bottom-up* approach, that uses reach-level survey measurements to form classes and infer higher-level controls; or a *top-down* approach, that uses higher-level controls to form classes and infer reach-level characteristics (Olden *et al.*, 2012). The approaches are also known as typologies of response or control respectively (Olden *et al.*, 2012).

Bottom-up typologies are often preferable as they take direct measurements of the feature of interest, whereas in top-down approaches features must be inferred. Bottom-up typologies rely on expensive and time-consuming survey data which may underrepresent certain areas and often focus on the immediate riparian environment rather than the whole catchment. The majority of applied typologies take a bottom-up approach by focusing on the reach and sub-reach levels (see review by Kondolf *et al.* 2003) leaving catchment level processes largely un-categorised.

Yet many classifications are hierarchical, with 19 out of 23 geomorphic channel classifications reviewed by Kondolf *et al.* (2003) including multiple levels. Of the 19 classifications that included multiple levels, only five included levels above reach-level (~10^1 m). Most management focused typologies at the reach-level (e.g. Rosgen, 1985; River Styles, Brierley and Fryirs, 2000; REFORM, Rinaldi *et al.*, 2016), are supplemented with GIS-derived characteristics of the survey reach but few also include wider catchment characteristics to better reflect the entire hierarchical framework. GIS-derived characteristics often reflect and upland-lowland gradient in river types (e.g. Jeffers, 1998), but there are other characteristics that influence rivers such as geology, climate and anthropogenic pressures in the catchment. There is therefore a need for top-down typologies that encompasses catchment controls to complement bottom-up approaches. As we explore here, advances in machine learning techniques may provide a means to improve the incorporation of variation and identification of natural boundaries in typology development.

Table 3.1. *Comparison of the number of local and catchment controls used to classify reaches and waterbodies (denoted by *) in previous typologies in Great Britain (an X indicates the corresponding control was included in the typology).*

	Local controls								**Catchment controls**						
	Site altitude	Site slope	Distance to source	Height of source	Channel geometry	Stream power	Discharge	Floodplain width	Elevation	Upstream area	Geology	Climate	Baseflow Index	Morphometry	Land cover
Jeffers (1998)	X	X	X	X											
Holmes et al. (1998)	X	X			X						X				
*UKTAG (2003)**									X	X	X				
*Acreman et al. (2008)**										X		X	X		
Bizzi and Lerner (2012)		X				X	X	X							
This typology*									**X**	**X**	**X**	**X**		**X**	**X**

3.2.1.2 Research design utilising national datasets and machine learning

Top-down typologies are built on continuous GIS-derived datasets for complete system coverage regionally, nationally, or even globally. Such typologies are useful for river management as there is no need for survey data and associated biases (see example of a top-down applied typology routinely used in river management by Acreman *et al.* (2008). Previous attempts at top-down typologies have been criticized for using a small number of variables relating to only few aspects of catchment functioning; for example, the current typology employed by the WFD, separates catchments based only on upstream area, elevation and geology (UKTAG, 2003) (**Table 3.1**). This causes overlap between river types because of external elements not included in the typology such as vegetation, climate and natural variability (Naura *et al.*, 2016). In particular, geomorphic characteristics of catchment morphometry that influence hydrological and sedimentological inputs to reaches (Schumm and Lichty, 1965) are often only accounted for via elevation (**Table 3.1**). Using few variables may thus result in poor distinction in river reach features between waterbody types (Naura *et al.*, 2003). Therefore, the typology developed here aims to capture a wider range of catchment controls that influence reach-level features than usually considered by existing typologies (**Table 3.1**).

A number of statistical techniques are available derive classifications from multivariate datasets (Liakos *et al.*, 2018), although many are hampered by the difficulty of separating individual controls on reach features because of the confounding effects of cross-correlation (often found between environmental variables; Feld *et al.*, 2016). To overcome this challenge here, the machine learning SOM method is selected because it can accommodate the non-parametric, categorical, and cross-correlated nature of the data available to characterise catchment controls (in contrast to other data reduction techniques, such as ordination). It also enables intuitive visual interpretation of gradients in catchment characteristics and other patterns hidden by the linearity of other methods. SOM is an unsupervised artificial neural network technique developed by Kohonen (1982) and has previously been used in river classifications of chemical and biological quality (Astel *et al.*, 2007; Walley *et al.*, 1999) and reach-level geomorphic drivers (Bizzi and Lerner, 2012). The SOM technique allows for a solely top-down typology to be developed at the national level, combining multiple catchment controls, including morphometric and anthropogenic characteristics for the first time in England and Wales (**Table 3.1**). To ensure the typology is useful for managers, the outputs from the SOM must be split into a practical number of catchment types (Kondolf *et al.*, 2003). The typology may have multiple uses, but in this study it is evaluated with survey data to explore evident linkages between catchment controls and reach response. The evaluation of the typology with survey data is a method used by other top-down approaches (Acreman *et al.*, 2008) and adds credibility to the typology.

3.2.2 Data and methods

The top-down typology of catchment controls was developed using multiple GIS-derived characteristics for waterbodies in England and Wales. The characterises were reduced using the SOM machine learning approach and the output was divided into a practical set of types, derived through hierarchical clustering, to determine typology classes. The functional applicability of the typology was evaluated using inferential statistics to determine whether reach-level features are distinguishable between waterbody types.

3.2.2.1 Catchment characteristics data

WFD waterbodies, sub-units of catchments, were used as the study unit for the typology. Waterbody boundaries are drawn when a river crosses an altitude, catchment area or dominant geology threshold, or at highly engineered or major tributaries (UKTAG, 2003). Coastal waterbodies were removed because of their tidal influence so only river waterbodies were included in the study (n=4485). Although the waterbody is a relatively coarse unit for classification and is not included in geomorphic hierarchical frameworks such as REFORM (Rinaldi *et al.*, 2016), it is a commonly used delineation of the landscape for extracting catchment controls, for example having previously been used to classify abstraction targets in the UK (Acreman *et al.*, 2008) (**Table 3.1**). Being sub-units, waterbodies do not capture the entire upstream area which may be very large (e.g. the Thames River Basin takes up ~16% of the surface area of England) but instead focus on catchment controls in a more localized landscape setting. Connectivity to upstream waterbodies is not directly considered but the cumulative catchment area characteristic indicates the position of the waterbody within the wider catchment (**Table 3.2**).

For each waterbody, 22 GIS-derived characteristics were extracted from continuous datasets to represent the morphometry, climate, geology and land cover of the waterbodies. Characteristics were summarized within each waterbody using ArcGIS v10.3 (**Table 3.2**). Multiple characteristics were used so that a range of influences on river functioning are captured by the typology. **Table 3.2** provides descriptions of how each catchment characteristic contributes to river functioning at the reach-level and the data and methods used to extract the characteristics using GIS are described below.

Morphometric catchment characteristics were calculated from the Centre for Ecology and Hydrology's (CEH) 50 x 50 m digital terrain model (CEH, no date-b; Morris and Flavin, 1990, 1994) for each waterbody using spatial analyst module in ArcGIS v10.3 following the methods indicated in **Table 3.2**. Maximum cumulative catchment area, the number of upstream grid cells flowing into an individual cell, was extracted for each waterbody (Morris and Flavin, 1990, 1994). The CEH's 1:50,000 blue-line network was used to calculate drainage density in each waterbody (CEH, no date-a; Moore *et al.*, 1994).

Rainfall characteristics were extracted from a 5 x 5 km grid of the number of days per month with over 1 mm precipitation (Met Office *et al.*, 2017; Perry and Hollis, 2005). Annual average was calculated as the mean of all months between 1961 and 2016. Seasonality of rainfall occurrence was extracted as the ratio of spring to winter mean rainfall with 1 indicating no seasonal rainfall and 0 indicating winter dominated rainfall. Mean annual average rainfall and seasonality were extracted for each waterbody.

Geology characteristics were obtained by simplifying the bedrock deposit map at 1:625,000 scale (BGS, no date) into broad geological classes following Harvey *et al.* (2008b), with four classes (hard rock geology, chalk, other limestone and sandstone) retained for analysis. Rocks considered to be major UK aquifers were also included following Vaughan *et al.* (2013). Land cover data was obtained from the CEH's 2007 land cover map at 25 x 25 m resolution (CEH *et al.*, 2014) and the six most prevalent land covers were retained for analysis. The percentage cover of each geological and land cover class within each waterbody was extracted using GIS. The characteristics were scaled and centred (i.e. converted to standardised z-scores) so all characteristics have equal importance during SOM training.

Table 3.2. *List of GIS-derived catchment characteristics used to create the typology and description of their control on river functioning. Units and source for the method is indicated where appropriate.*

Catchment characteristic	Units	Control on river functioning
Morphometry		• Area (related to discharge;(Knighton, 1998)and slope drive stream power which is related to sediment transport and sorting (Rice and Church, 1998). • Elevation, standard deviation of elevation and TPI (Weiss, 2001) reflect topographic variability, erosivity and therefore sediment availability. • Dissected catchments with high drainage density and roughness (TPI) have greater channel heterogeneity (Benda *et al.*, 2004b). • TWI (slope's ability to evacuate upstream water; Beven and Kirkby, 1979) and HI (whether hillslope or fluvial processes are dominant; Willgoose and Hancock, 1998) reflect dominant geomorphic processes. • Catchment shape (circularity ratio; Miller, 1953) reflects hydrograph magnitude and time to peak (Gregory and Walling, 1973).
Cumulative catchment area	km^2	
Mean slope	deg.	
Mean elevation	m	
Standard deviation elevation	m	
Topographic Position Index (TPI)	0-1	
Topographic Wetness Index (TWI)	0-1	
Drainage density	km/km^2	
Hypsometric Index (HI)	0-1	
Circularity ratio	0-1	
Climate		• Rainfall volume influences the magnitude and duration of flood peak (Singh, 1997). • Rainfall seasonality determines runoff intensification during floods (Flores *et al.*, 2006)
Mean annual number of days with rain >1mm	n	
Seasonal rainfall ratio	0-1	
Geology		• Rock permeability influences the flashiness of the hydrograph (Holmes *et al.*, 2002; Sear *et al.*, 1999). • Rock type determines the sediment calibres available in the catchment (Naura *et al.*, 2016)
Hard rock	%	
Other limestone	%	
Sandstone	%	
Chalk	%	
Aquifer	%	
Land cover		• Wooded catchments and unmodified floodplain store water and release it slowly whereas impermeable surfaces and highly connected drainage network in urban and arable areas increase flood peaks (Dadson *et al.*, 2017). • Arable land practices are related to increases in fine sediments in channels (Wharton *et al.*, 2017). • River management works in urban and arable areas (such as dredging and straightening) increase channel dimensions creating depositional, homogeneous reaches (Sear *et al.*, 2003).
Woodland	%	
Improved grassland	%	
Semi-natural grassland	%	
Mountain, heath, bog	%	
Arable	%	
Urban	%	

3.2.2.2 Self-organising maps (SOMs)

SOMs display the signal from high-dimensional data onto a low-dimensional network. SOMs are a black box technique, so utility is in holistic visual interpretation of the low-dimensional output rather than understanding underlying processes. In broad terms, the output layer (i.e. the self-organised map itself) contains neurons organised on a rectangular or hexagonal lattice grid to represent the entire dataset (in this case hexagonal grid was chosen because it does not favour horizontal or vertical direction (Kohonen, 2001)). The user determines the dimensions of the grid from the ratio between the greatest two eigenvalues of the input variables (Park *et al.*, 2006). Actual height and width are set to return the number of cells closest to $5\sqrt{N}$ where N is the number of samples (Vesanto, 2000), in this case N=4485 waterbodies. Therefore, a grid with dimensions of 12 x 28 cells is established, to produce a total of 336 cells.

Each neuron (or grid cell) has an n-dimensional weighting vector, in this case n=22, the number of catchment characteristics (**Table 3.2**). The neurons are related to neighbouring neurons which defines the map's topology. For each iteration in the SOM training algorithm, a sample (in this case, a waterbody) is selected at random and the distance in data space between it and all the weight vectors is calculated. The algorithm optimises the weight vectors at each iteration step. The output grid therefore comprises cells containing similar waterbodies which are mapped closely to other cells with similar characteristics on the grid. The output can be visually interpreted as a number of heatmaps for each characteristic and the unified distance matrix (U-matrix) indicating the distance between neighbouring cells. The SOM analysis was conducted in the 'kohonen' v3.0.7 package (Wehrens and Kruisselbrink, 2018) in R v3.5.1 (R Core Team, 2018), with code for analysis available online [doi.org/10.5281/zenodo.3558120].

3.2.2.3 Cluster analysis

Hierarchical clustering was then performed on the SOM output grid to delineate clusters of similar waterbody types. This is a 'natural' method of classification, as opposed to 'special' classification in which arbitrary lines are drawn across a continuum. Special classification has often been applied, for example the River Habitat Survey classification (Jeffers, 1998a) and the current WFD System A typology (UKTAG, 2003), but is highly criticised (Wright *et al.*, 1984). In contrast, as a natural classification approach, hierarchical clustering identifies latent thresholds in the data to group inherently similar objects together. The optimal number of clusters was determined using the Davies-Bouldin index (Davies and Bouldin, 1979) where the lowest values represent small within-cluster scatter and good separation between clusters. This index has been used by multiple studies to determine the optimum number of clusters for an SOM output (e.g. Astel *et al.*, 2007; Bizzi and Lerner, 2012). However, expert judgement based on knowledge of the system is also required when determining whether the number of clusters is fit for purpose (Kondolf *et al.*, 2003).

3.2.2.4 Evaluating the typology with River Habitat Surveys

To test the applicability of the waterbody typology to reach-level habitat features, data collected as part of the national River Habitat Survey monitoring programme (RHS; Raven *et al.*, 1996) was utilised. RHS is a standard methodology for hydromorphological assessment under the WFD (CEN, 2004) collected by England's Environment Agency, with over 24,000 sites sampled since 1994, observing over 100 river habitat features with every 500m survey reach. While the detail of river processes recorded in the survey is limited (Belletti *et al.*, 2015), the wide spatial and temporal coverage of this dataset means that it has been used to create numerous bottom-up typologies (Bizzi and Lerner, 2012; Harvey *et al.*, 2008a; Jeffers, 1998a; Vaughan *et al.*, 2013) and makes it a useful means of validating this top-down typology. RHS surveys were not sampled with the intention of being used with waterbodies, which means that the number and distribution of RHS sites within waterbodies varies. Therefore, we expect there to be variation in habitats within waterbodies due to local controls.

Six habitat indices were calculated from the RHS observations for use in this study (**Table 3.3**); two summary indices and four individual indices. The summary indices – Habitat Quality Assessment (HQA), a measure of diversity and naturalness, and Habitat Modification Score (HMS), a measure of anthropogenic modification – were calculated using scores for individual features weighted by expert opinion (see Raven *et al.*, 1998 for details). HQA and HMS are semi-quantitative measures of reach condition but are regularly used for river quality assessment.

The remaining four indices were calculated directly from individual RHS observations to reflect physical habitat conditions at each site. Reach averaged sediment size and flow type speed were estimated using methods used in previous studies (Davenport *et al.*, 2004; Emery *et al.*, 2004; Harvey *et al.*, 2008b). The sediment size and flow type speed indices were inverted so the highest values indicate coarser sediment and faster flow respectively. Sediment size and flow type speed diversity were also calculated for each site using Simpson's diversity index (Simpson, 1949).

To test if the waterbody typology reflected habitat conditions in reaches, the distribution of habitat indices values from all the RHS sites located in each waterbody type were compared. A Kruskal-Wallis test, followed by Dunn post-hoc test with False Discovery Rate correction (Benjamini and Hochberg, 1995) to the p-value, were conducted to test the significance of differences in habitat indices between waterbody types.

Table 3.3. *Habitat indices calculated from the national RHS dataset used to evaluate the typology and the ranges of the indices.*

Habitat index	Mean Scores (Range)
Summary indices *Overview of reach condition for river quality assessment*	
Habitat Quality Assessment (HQA)	42 (1-94)
Habitat Modification Score (HMS)	1055 (1-7715)
Individual habitat indices *Quantify individual components of reach condition that reflect physical habitat*	
Flow type diversity	0.39 (0-0.84)
Sediment diversity	0.30 (0-0.82)
Flow type speed	3.29 (0-7.9)
Sediment size	2.46 (-9-8)

3.2.3 Results

The SOM analysis produced heatmaps that capture gradients in catchment controls that were then sub-divided into seven waterbody types through hierarchical clustering. The characteristics of each type and the spatial distribution of types across England and Wales were assessed before the typology was evaluated against reach-level survey data.

3.2.3.1 Interpreting SOM outputs

The SOM output was assessed using several measures (**Figure 3.1**) overlain on the same grid. The grid represents the topological configuration of the waterbodies based on their catchment characteristics, where each grid cell contains several waterbodies (between 1 and 34 waterbodies) with similar characteristics (**Figure 3.1a**). The topological configuration of the map means that waterbodies in each grid cell are most similar to those in neighbouring grid cells, depicted by the U-matrix in **Figure 3.1b**, where low values indicate that the grid cell is similar to neighbouring grid cells.

Hierarchical clustering was applied to the SOM output to identify typology classes. The decision of which number of classes to use depends on the intended purpose, as successful typologies must be interpretable to be fit for purpose (Kondolf *et al.*, 2003). Here seven clusters were selected based on the Davies-Bouldin index, a statistical measure of clustering quality, and because seven clusters sufficiently captured the complexity of catchment characteristics that influence river functioning whilst remaining interpretable (see **Appendix 3A** for further discussion relating to the number of clusters chosen).

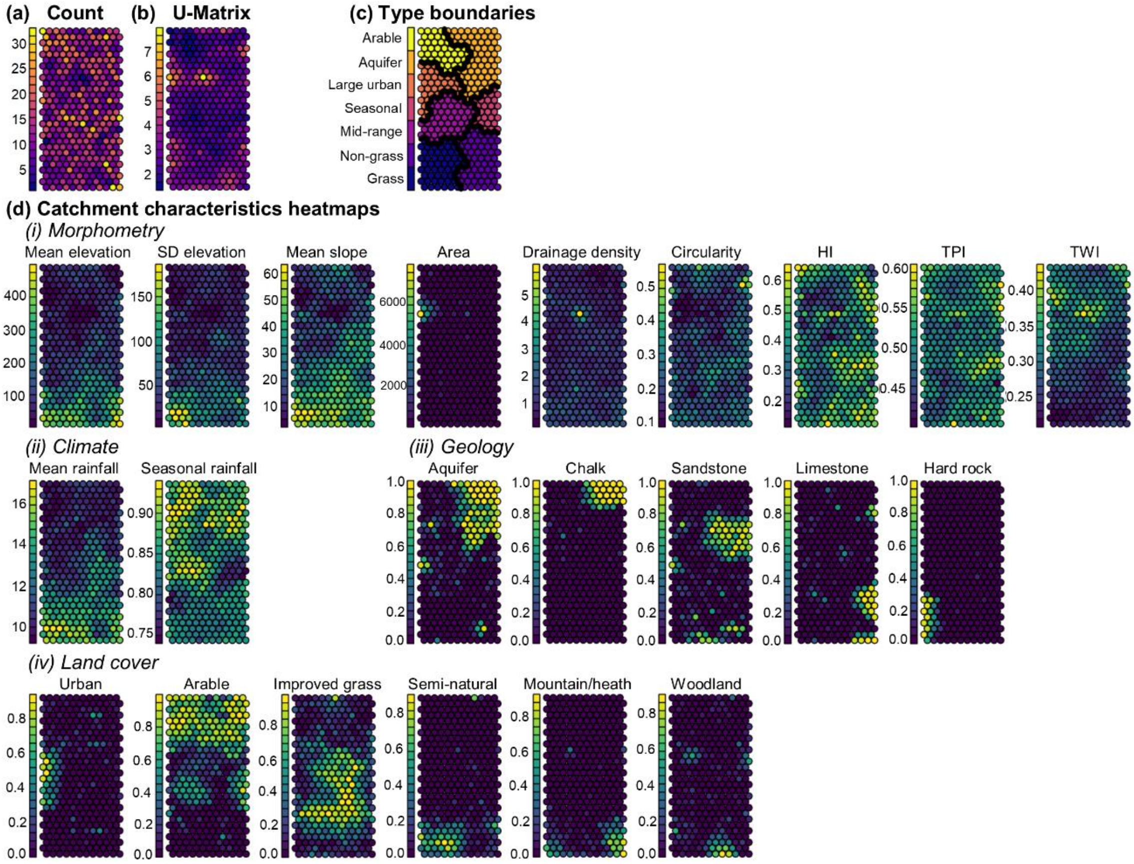


Figure 3.1. *SOM output grids: (a) the number of waterbodies within each grid cell; (b) U-matrix (unified distance matrix) indicating the difference between neighbouring grid cells; (c) catchment type boundaries identified from the hierarchical clustering analysis. The name attributed to each type is described in the text; (d) heatmaps of characteristics displayed on the SOM grid (scale bars in units of each characteristic shown in **Table 3.2**).*

The final waterbody type boundaries are presented in **Figure 3.1c** for comparison with the SOM heatmaps (**Figure 3.1d**). The heatmaps show the distribution of values for each morphometric, climatic, geological and land cover characteristic across the SOM grid (**Figure 3.1d**). They indicated a gradient from upland to lowland waterbodies, from the bottom to the top of the heatmaps. At the upland end of the gradients there was higher elevation, slope and rainfall, greater run-off (indicated by TWI), drainage density, seasonal rainfall, harder geologies and more natural land covers, and vice versa for the lowland end of the gradient.

Further inspection of the heatmaps indicated additional patterns and anomalies. The morphometric characteristics HI, TPI and circularity showed high levels of variation indicating differing degrees of roughness and catchment development (Willgoose and Hancock, 1998) across the upland-to-lowland gradient. There was also a secondary gradient from waterbodies with homogeneous to heterogeneous landscapes running from the left to right-hand side of the heatmaps with higher HI, TPI, circularity, slope and rainfall values on the right. Other anomalies such as extreme high drainage density values that to not sit in the gradient were apparent, along with a group of waterbodies with high percentage urban land cover and high cumulative catchment area on the left-hand side. Differences in the middle of the upland-lowland gradient were also shown in improved grassland land cover and highly seasonal rainfall.

3.2.3.2 The waterbody typology

The boundaries of the seven selected waterbody types are displayed in **Figure 3.1c** in relation to their catchment characteristics and are named based on the interpretation of the authors. The typology was mapped across England and Wales in **Figure 3.2a**. The seven types fit into three broader categories – upland, midland and lowland – based on the dominant upland-lowland gradient displayed in the heatmaps in **Figure 3.1d**.

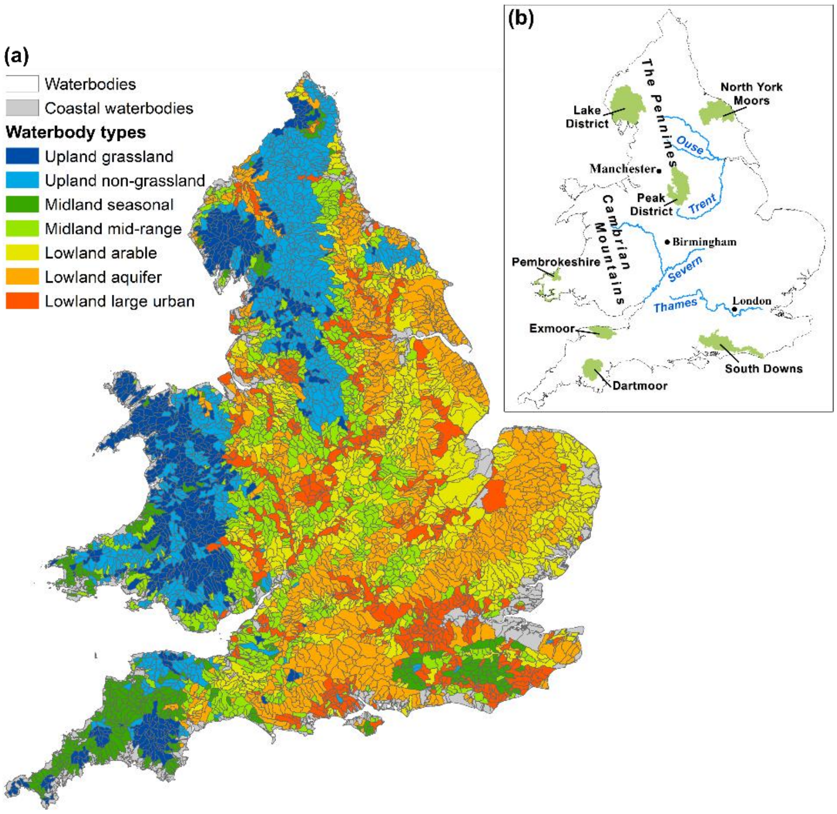


Figure 3.2. (a) Map of catchment typology for England and Wales based on the SOM analysis with the names attributed to each type. (b) Location of features in England and Wales that are mentioned in the text (for readers unfamiliar with the geography of England and Wales); green areas indicate national parks (Office for National Statistics, 2016).

Upland waterbody types

Upland waterbody types were defined by high elevation (over 350m), slope (over 50 degrees) and rainfall (over 14 days with >1 mm rainfall a year) (**Figure 3.1d**). Both upland types exhibited high U-Matrix values (**Figure 3.1b**) indicating that waterbodies within upland waterbodies are diverse within this overall gradient.

Upland grassland types (n=608) were distinguished as having the highest slope and standard deviation of elevation values, lowest TWI and are dominated by natural grassland and hard rock geology (**Figure 3.1d**). This suggests deep valleys in a steep impermeable landscape with high levels of runoff. This type is predominantly located in the Lake District, Cambrian Mountains and Dartmoor (**Figure 3.2**).

Upland non-grassland types (n=824) had higher circularity, HI and TPI values (**Figure 3.1d**) indicating a more rugged, heterogeneous landscape dominated by hillslope processes (Willgoose and Hancock, 1998). This type had limestone geology and mountainous, heath, bog and woodland land covers and was located in the Pennines, North York moors and Exmoor (**Figure 3.2**).

Midland waterbody types

Midland types were more internally homogeneous than upland or lowland types (**Figure 3.1b**). Both midland types had similar mean elevations (~150-250 m) and were dominated by similar geologies, improved grassland and arable landcovers. Differences were primarily in the morphometric and climatic characteristics (**Figure 3.1d**).

Midland seasonal types type (n=351) had highly seasonal rainfall with higher slopes, rainfall, circularity, HI and TPI compared to mid-range types (**Figure 3.1d**). Seasonal waterbodies were the least numerous, limited to the South Downs, the South West and Pembrokeshire (**Figure 3.2**).

Midland mid-range types (n=732) had lower slopes and were less rugged landscapes. They had less rainfall which was less seasonal. This type had a wide spatial distribution often adjacent to upland types or representing comparatively upland areas in central England (**Figure 3.2**).

Lowland waterbody types

Lowland types had lower elevation, slope and rainfall than other types. Lowland arable types (n=681) had the lowest elevation and rainfall. They were dominated by arable land covers (~80% cover) and high TWI indicating low floodplain locations. There was little variation in catchment characteristics within this type (**Figure 3.1b**). Arable types were evenly distributed across the country in the floodplain areas of major rivers and dry, low-lying areas on the east coast (**Figure 3.2**).

Aquifer types (n=892) are had more diversity within the class than arable waterbodies (**Figure 3.1b**), despite also being dominated by arable land. This is likely because the class boundary reflected the aquifer boundary that contained both chalk and sandstone permeable geologies. Aquifer types had low drainage density with a slightly rougher terrain than other lowland classes, indicated by higher slopes, HI, TPI and circularity (**Figure 3.1d**). The distribution of aquifer waterbodies followed bands of permeable geology across England (**Figure 3.2**).

Large urban types (n=397) were distinguished by their high percentage of urban land cover (>50%) and large cumulative catchment area, indicating that they are downstream waterbodies. The boundary of this type extended towards the upland end of the heatmap, indicating that large urban conditions occur over a range of mid-low elevations and conditions. This is likely why there is higher heterogeneity of characteristics within this category than others (**Figure 3.1b**). Large urban waterbodies were centred around large urban settlements such as London, Birmingham and Manchester or large main rivers such as the Ouse, Trent, Severn and Thames etc. (**Figure 3.2**).

3.2.3.3 River habitat differentiation between types

Reach-level characteristics were compared between the seven waterbody types to evaluate whether the summary indices of reach quality and individual physical habitat indices (**Table 3.3**) vary between types. All six river habitat indices showed a range of significant differences among waterbody types using the Kruskal-Wallis test ($p<0.01$). The Dunn post-hoc test indicated that most waterbody types had significantly different indices from one another ($p<0.05$; **Figure 3.3**).

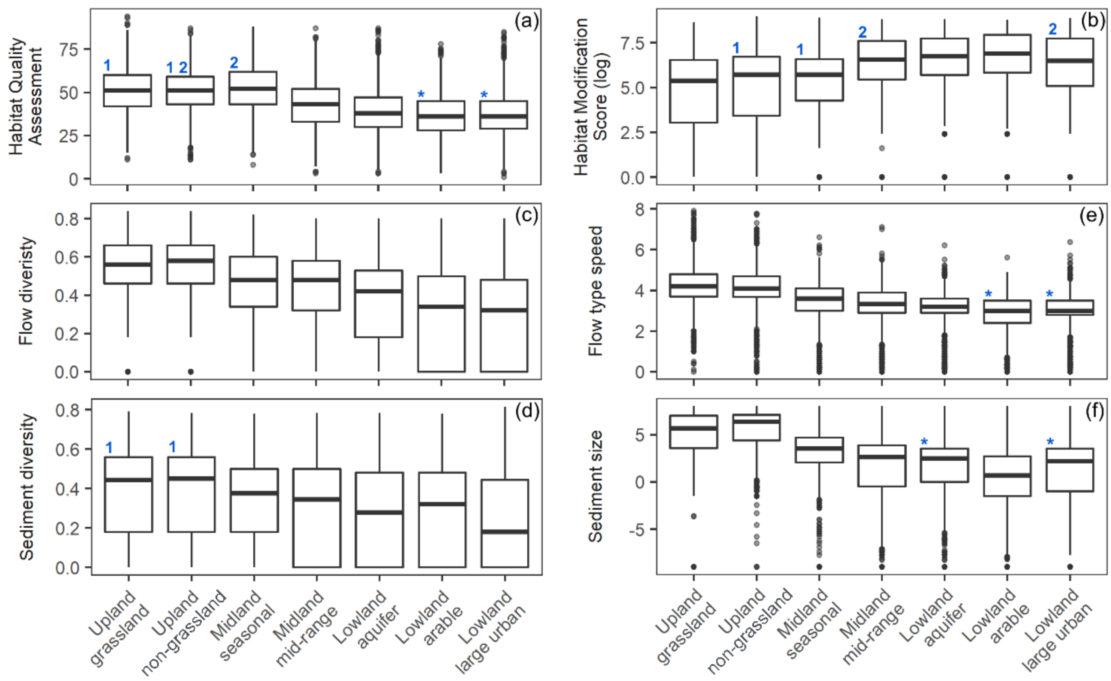


Figure 3.3. *RHS variable distributions for each catchment type (HMS plotted on a log-scale). Types with no significant difference (p>0.05) between each other, as a result of the Dunn test, are indicated by numbers. *Indicates distributions with a significant difference of p<0.05, all other differences p<0.01.*

Flow type speed, sediment size and flow type diversity differed significantly between all types (**Figure 3.3c, 3.3e** and **3.3f**). Their distributions predominantly reflected the upland-lowland gradient in waterbody types, with coarser sediments and faster and more diverse flow types in upland waterbody types. Lowland arable waterbodies tended to have the lowest index values of the three lowland types for these indices.

Sediment diversity also exhibited an upland-lowland trend although there are no significant differences in diversity between the two upland classes (**Figure 3.3d**). Sediment diversity values were lowest in large urban waterbodies despite lowland arable types exhibiting lower sediment sizes (**Figure 3.3f**).

For both flow indices (**Figure 3.3c** and **3.3e**), there was a steady decline in index value through the waterbody types. For sediment indices, there was a larger difference between seasonal and mid-range types that was less evident in the flow indices (**Figure 3d** and **3f**). Sediment size was also greater in upland non-grassland than upland grassland waterbodies (**Figure 3.3f**).

The summary indices, HQA and HMS (**Figure 3.3a** and **3.3b**), also reflected the upland-to-lowland gradient with high habitat quality and low modification scores in upland sites compared to lowland sites. There were more similarities in summary indices between waterbody types than for the individual habitat indices. HQA was not significantly different between the upland grassland, upland non-grassland or midland seasonal types and HMS was not significantly different between midland mid-range and lowland large urban waterbodies, with lowland arable waterbodies exhibiting the greatest modification scores (**Figure 3.3b**).

While there were many statistically significant differences between waterbody types, **Figure 3.3** also highlights the broad range of river habitat indices within each type.

3.2.4 Discussion

3.2.4.1 A practical and applicable typology of catchment controls for waterbodies in England and Wales

Selected catchment controls have been used in previous applied typologies to delineate homogeneous river sections (Brierley and Fryirs, 2000; Rinaldi *et al.*, 2016) but the associations between catchment controls, and the response of river reaches to their combined effects, is often not considered. The typology presented here is less focused on classifying reach processes for local management than previous typologies. Instead, the typology was designed to capture multiple catchment controls and their associations for identifying natural boundaries in catchment functioning for strategic management at the national level.

The typology of catchment controls developed using the SOM approach for waterbodies in England and Wales was successful at differentiating between key features of the landscape including national reserves, topographical and geological features, major rivers and urban centres (**Figure 3.2**). The approach incorporates multiple catchment characteristics that have a functional control on river reaches (**Table 3.3**) rather than being limited to only characteristics that are not correlated with one another. Furthermore, the typology boundaries are based on naturally occurring thresholds in the data identified by the clustering algorithm rather than arbitrary boundaries.

These factors likely explain why this waterbody typology differentiates habitat features between types better than the current WFD System A typology. When evaluated against flow type, substrate size and geomorphic activity indices derived from semi-natural RHS sites, 0% of WFD System A types were statistically different to all the other types (at a significance level of $p<0.05$; Naura *et al.*, 2003). However, in this typology, using the same level of significance, up to 100% of types produced statistical differences in habitat indices between all other types (**Figure 3.3**), including 42-57% for the summary indices used to assess the quality of reaches. This indicates that this typology has relevance for river managers and conceptually improves upon the current WFD System A typology, which is based solely on elevation, catchment area and geology (**Table 3.2**) and has arbitrary boundaries between categories (UKTAG, 2003).

The strength of this typology is the range of catchment characteristics included that often showed cross-correlations (**Figure 3.1d**). Cross-correlation makes it difficult to isolate individual effects from catchment controls as they interact (Feld *et al.*, 2016). This is because catchment controls are not independent (Schumm and Lichty, 1965) and therefore grouping waterbodies with similar controls is beneficial rather than relying on a single control to describe all catchment influences.

The inclusion of multiple characteristics was possible due to the adoption of the SOM method. This and other machine learning techniques are becoming more prevalent in multivariate analysis as they can deal with natural artefacts of many environmental datasets which often make multivariate environmental analyses challenging (Feld *et al.*, 2016). The heatmap outputs from the SOM (**Figure 3.1d**) also allow for easy visualisation of variable distributions, positive and negative correlations between variables such as the upland-lowland gradient, and anomalies such as the higher drainage density anomaly in the large urban type (ASCE, 2000; Astel *et al.*, 2007).

3.2.4.2 Critique of the typology

Whilst the waterbody typology shows promising differentiation between landscape (**Figure 3.2**) and reach features (**Figure 3.3**), its limitations must be understood to ensure it is not applied for management in ways that are inappropriate given its design. The most obvious example of limitations is the wide ranges of habitat index values within each waterbody type, despite overall significant differences between most types (**Figure 3.3**). As the aim of this paper was to create a waterbody typology that can be applied widely, this is expected, but reasons for these variations are discussed below to highlight limitations of the typology.

The variation in characteristics within waterbody types is greatest in aquifer, large urban and both upland types (**Figure 3.1b**). Creating more types may capture more variation and the selection of the number of types in any typology is ultimately subjective (Bizzi and Lerner, 2012; Tadaki *et al.*, 2014), but is aided by statistical measures and expert opinion (for the methods used here, see **Appendix 3A**). An interpretable classification will never capture the whole range of variation of its population, nor is it expected to, but it must capture enough variation to be fit for purpose. As discussed above, we believe that seven types are appropriate to capture the variation in catchment controls at this national level, evidenced by evaluating the types against survey data (**Figure 3.3**).

The limitations of the RHS dataset, used here to represent reach features, should also be noted. The RHS was not designed as a geomorphological survey to capture dynamic process (Newson *et al.*, 1998a) but does include the presence/absence of features that are useful to estimate dominant channel habitat conditions over a standardised 500m reach. The identification of dominant features present at each transect in the survey means that the diverse conditions of the reach may be underestimated which may mute more extreme differences between waterbody types. However, although the RHS is not detailed, it does provide a wide spatial coverage with a consistent methodology that makes it a valuable tool for use in national typologies (Holmes *et al.*, 1998; Jeffers, 1998a).

The waterbodies used as the unit for the typology developed here, are much larger than reach or sub-reach units employed by bottom-up typologies (e.g. Brierley and Fryirs, 2000; Rinaldi *et al.*, 2016; Rosgen, 1985), which has practical benefits. For example, the resolution of the GIS-derived datasets used to build the typology can be relatively coarse and there are numerous RHS surveys available within each waterbody type to effectively evaluate the typology. The waterbody unit also reflects policy units that are widely applied in river management in Europe (European Commission, 2000) Providing a continuous typology across the landscape is not possible if one is relying on survey data alone. However, the use of waterbodies as sub-units of the wider catchment means that controls from upstream of the waterbody are not considered. Only the cumulative catchment area characteristic indicates the position of the waterbody within the wider catchment which contributed to the large urban waterbody type, separating waterbodies at the downstream end of catchments from other waterbody types. The use of a relatively large study unit also means that variation will be present within types because each waterbody contains a range of processes and local pressures such as sediment mining, dams and channelization that are not included in the typology (a limitation of this methodology). The aim of this typology however, was to capture the catchment controls that influence the reach, rather than directly classifying reach processes and features such as channel stream power, slope, and planform, which have been the focus of previous top-down and bottom-up typologies (e.g. Bizzi and Lerner, 2012; Brierley and Fryirs, 2000; Rosgen, 1985). For increased utility of this typology for operational river management at a more local level, data on controls and characteristics at the reach-level should be integrated into the waterbody typology.

The typology also is a temporary snapshot of catchment controls, which is often a critique of river typologies (Kondolf *et al.*, 2003). While many catchment characteristics change over long timescales, such as morphometry or geology (~10^2 to 10^4 years), some characteristics are more temporally dynamic such as land cover and rainfall patterns (~10^1 to 10^2 years; Gurnell *et al.*, 2016). This is addressed to some extent by taking a long-term average of rainfall (from 1961 to 2016) and a land cover map for the time period most relevant to the validation surveys (2007). While this is not ideal, the top-down nature of this approach means the typology can easily be updated at a relatively inexpensive cost to the user as, and when, major landscape alterations are made or when new data become available. The typology is also evaluated with RHS surveys occurring over a long time period (1994 to 2015) each providing a snapshot of river features that change ~10^{-1} to 10^1 years rather than the long-term changes of the catchment controls. Although the link between catchment changes and channel features is complex, the fact the typology performs well when evaluated against over 20 years' worth of surveys suggests that the typology is relevant over long time periods.

Whilst there are limitations, primarily as a result of the selection of the top-down approach, the validation of the waterbody typology with reach-level data not only creates a useful typology tool with distinctive classes but enhances understanding catchment controls on reach habitats. The top-down method means that this approach can be applied to any waterbody with available data, without expensive and systematically biased surveys. However, the broad distribution of habitat features within each type (despite statistically significant differences; **Figure 3.3**) emphasises that this typology is not a substitute for detailed surveys and monitoring, but a means of assessing the spatial distribution of catchment controls at a national level. Future work should compare different datasets that reflect other aspects of the geomorphology or ecology of the channel to this typology.

3.2.4.3 Gradients and anomalies in catchment types and reach responses

The waterbody types show distinctive distributions of catchment controls reflecting dominant upland-lowland and secondary topographic heterogeneity gradients. Anthropogenic controls often follow these gradients but can occur independently. The response of habitat indices to the waterbody types reflects the gradients observed in the catchment controls.

Upland-lowland gradient

Many bottom-up typologies derived from RHS data detect a regional upland-lowland gradient using elevation and distance in the network (Jeffers, 1998a). In addition, others also found factors such as geology, climate and mean catchment slope to be useful descriptors of regional river habitat patterns (Harvey *et al.*, 2008b; Holmes *et al.*, 1998; Naura *et al.*, 2016; Vaughan *et al.*, 2013). Those that considered anthropogenic catchment pressures found them to only have a weak effect on habitat features (Jusik *et al.*, 2015; Naura *et al.*, 2016). We also observe an upland-lowland gradient present across morphometric, climatic, geological and anthropogenic catchment characteristics of England and Wales (**Figure 3.1d**; **Figure 3.2**), which justifies the validity of a multivariate typology.

The upland-lowland gradient across most characteristics is because of dependency between catchment characteristics that dictates the discharge of water and sediment to the channel (Schumm and Lichty, 1965) altering physical habitat features (Bunn and Arthington, 2002). The results indicate upland to lowland variation in a variety of processes that are strongly related to geology and topography, including reductions in sediment transport capacity, lower magnitude and frequency hydrographs and, perhaps most importantly, increasing anthropogenic pressures from upland to lowland waterbodies (Raven *et al.*, 2010). This is reflected in the habitat indices which decrease in habitat condition from upland to lowland (**Figure 3.3**). The distinct separation of habitat indices between each waterbody type, including the midland types, highlights the need to consider rivers along a gradient and not just upland or lowland polarisations.

Heterogeneity gradient

While the upland-lowland gradient is dominant both in explaining patterns of catchment characteristics (**Figure 3.1d**), and habitat indices distributions (**Figure 3.3**), a secondary gradient is identified in this waterbody typology. It is a gradient of topographic heterogeneity, driven by patterns in HI, TPI, land cover and geology. Previous studies identified an energy gradient within catchments, from upstream to downstream, as a secondary gradient (Jeffers, 1998a; Vaughan *et al.*, 2013). The distribution of energy within catchments is widely considered a key factor in distributions of geomorphological forms and processes (Church, 2002) and ecological communities (Gurnell *et al.*, 2010; Vannote *et al.*, 1980). However, this typology at a broader spatial level so internal waterbody variations are not accounted for. This emphasises the heterogeneity gradient that has not before been identified nationally. It shows that fluvial processes vary at the same point along the upland-lowland gradient as a result of landscape heterogeneity.

The heterogeneity gradient is related to energy, reflecting regional patterns of process. Heterogeneous waterbody types are more circular indicating flashier hydrographs (Gregory and Walling, 1973), have greater local ruggedness indicating greater coupling to hillslopes and flood responses (Church, 2002; Weiss, 2001) and greater hypsometric integrals suggesting greater dominance of hillslope processes (Willgoose and Hancock, 1998) than their counterparts at the same point in the upland-lowland gradient (**Figure 3.1d**). These morphometric variables are dependent on climate and geology (Schumm and Lichty, 1965), which create deviations from the upland-lowland gradient, such as higher elevation landscapes in lowland waterbody types due to the permeable geology, more easily eroded landscapes in upland limestone waterbodies and more seasonal rainfall producing flashier flood hydrographs in some midland waterbodies (Flores *et al.*, 2006). The permeable geology and natural, diverse land covers may also stabilise the hydrograph (Holmes *et al.*, 2002) creating a complex range of processes that are less prominent in the homogeneous waterbody types that are dominated by fluvial processes and anthropogenic land covers.

Catchments with a more variable topography are predicted to produce reaches with greater geomorphic heterogeneity (Benda *et al.*, 2004b). We also observe this as heterogeneous waterbody types tend to exhibit better habitat condition than their counterparts at the same point in the upland-lowland gradient (**Figure 3.3**). Others have also observed differences at similar points along the upland-lowland gradient; Holmes *et al.* (1998) found different macrophyte species at similar elevations which they attribute to geological differences. However, the heterogeneity gradient better explains the processes that influence reaches which are as a result of driving variables such as geology and climate. This highlights the utility of using multiple catchment characteristics, particularly morphometry, when exploring catchment controls opposed to solely measures of the upland-lowland gradient which do not capture the range of processes occurring regionally at similar elevations (**Figure 3.1d**).

Anthropogenic consistencies and anomalies

Integrated catchment management often focuses on anthropogenic controls, particularly pressures from agricultural and urban land (Downs *et al.*, 1991), but anthropogenic activity may be hard to distinguish from the upland-lowland gradient (Allan, 2004), as arable land dominates in lowland waterbodies (**Figure 3.1d**). Urban land cover crosses a range of low-mid elevations suggesting partial independence from the upland-lowland gradient, although it is less dominant in upland rural regions (Raven *et al.*, 2010). Large urban types are however are located at the homogeneous end of the heterogeneity gradient, likely because of limited topographic variability and the location of urban centres in large floodplains dominated by fluvial processes (**Figure 3.1d**).

While anthropogenic land covers reflect gradients in more natural catchment characteristics, habitat indices vary between waterbody types dominated by these land covers. In some cases, habitat indices reflect this gradient, for example, aquifer waterbodies which are dominated by arable land cover but are heterogeneous, frequently has higher habitat indices than other lowland waterbodies (**Figure 3.3**). This was also reported in Holmes *et al.*'s (1998) macrophyte typology and is expected as groundwater streams are often characterised by their gravel beds, moderate flow and relatively steep gradient (Berrie, 1992).

In contrast, lowland arable types frequently have the finest sediments (**Figure 3.3f**), expected partly because of sediment fining associated with the upland-lowland gradient (Naura *et al.*, 2016), but also because of increases in fine sediment from agricultural practices (Wharton *et al.*, 2017) and the widening and deepening of agricultural drainage ditches that create depositional environments (Poff *et al.*, 1997). Arable type waterbodies also have the highest modification score which follows the upland-lowland gradient but is surprising as large urban waterbodies commonly have modifications for flood and erosion protection (Sear *et al.*, 2003; Walsh *et al.*, 2005). Yet, large urban waterbodies have the lowest diversity scores (**Figure 3.3c** and **3.3d**), often with homogeneous flow and sediments, because of management practices such as over-widening, straightening and dredging for flood protection in urban centres (Sear *et al.*, 2003). It is therefore critical to consider anthropogenic catchment controls in the context of wider catchment processes as they may exaggerate or resist underlying natural gradients.

3.2.5 Conclusions

The typology developed and presented here is designed to reflect multiple catchment controls on river reaches, a development on previous typologies that classify reach features using survey data and only consider a subset of possible catchment controls. The use of SOMs combined with hierarchical clustering on this wide range of catchment characteristics has produced a national-level waterbody typology map for 4,485 waterbodies in England and Wales.

The typology shows clear differentiation of key landscape features – such as urban centres, national parks, geological features and topographic gradients – and river habitat indices extracted from the RHS dataset. The typology was evaluated with survey data and found to have functional significance, making it valuable for understanding catchment controls on reach features that are important to river managers. The top-down approach utilising solely GIS-derived data allows the typology to be continuous and easily revised as datasets are updated. The same methodology can be applied to other countries with available GIS data and monitoring data for validation. It is therefore clear that top-down approaches can be useful in river typologies, allowing the controls on rivers to be classified rather than just the responses to provide an additional layer of understanding.

The typology map in **Figure 3.2** may provide a useful tool for useful assessment of catchment controls in waterbodies, including the type of characteristics that may be influencing the river systems and broad habitat conditions. It can be rapidly applied without the need for time-consuming or expensive surveys to assess the spatial distribution of catchment controls at a national level to aid more strategic management. Integration with more localized data is also possible and would increase the utility of the typology from an operational perspective to river management. Although it is not a substitute for detailed surveys and monitoring, the use of. field surveys in conjunction with this broad representation of functional catchment controls should enable for a holistic assessment of catchment controls on river reaches. This may discourage a 'one-size fits all' approach to river management and offer a step towards better integrated catchment management.

Author Contributions: Conceptualization, E.H. and N.C.; methodology, E.H., J.M. and M.C.; formal analysis, E.H.; investigation, E.H.; writing—original draft preparation, E.H.; writing—review and editing, E.H., J.M., N.C. and M.C..; visualization, E.H.; funding acquisition, E.H.

Funding: This research was funded by Natural Environmental Research Council (NERC), grant number NE/L002485/1. E.H. is the recipient of the funding.

Acknowledgments: The authors would like to thank the Environment Agency and the Centre for Ecology and Hydrology for access to the data used in this paper. The River Habitat Survey data and WFD waterbody boundaries can be access through the UK government portal (data.gov.uk). Links to all other datasets used provided in the relevant reference.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

3.3 SUPPLEMENTARY WORK

This section details work that was not included in the paper in Section 3.2 for reasons of space and focus. It includes a comparison of the SOM machine learning method used to derive the typology in the paper, to another more traditional statistical method, Hierarchical Clustering on Principal Components (HCPC) (Section 3.3.1). Then, the waterbody typology described in the paper is used to explore diversity in catchment-level effects within the entire catchment as a potential application of the typology for scientific research (Section 3.3.2).

3.3.1 Comparison of SOM-derived typology to HCPC-derived typology

Before the SOM machine learning approach, the HCPC method was applied to the task of data reduction of the 22 catchment characteristics and cluster delineation. HCPC was originally chosen as it combined Principal Components Analysis (PCA), a method that has previously been used for data reduction prior to typology creation (Jeffers, 1998a), with clustering algorithms to identify natural boundaries between types opposed to arbitrary typology boundaries (e.g. Jeffers, 1998), a key benefit of this statistical approach (Azam *et al.*, 2017). While SOM, PCA and hierarchical clustering have been compared individually (Astel *et al.*, 2007), the combined HCPC approach has not been compared to the SOM approach, therefore this comparison is presented here as a means of justification for the selection of the SOM approach.

3.3.1.1 The HCPC method

HCPC combines PCA, hierarchical clustering and partitional (k-means) clustering (Husson *et al.*, 2010). HCPC has been used for bottom-up catchment typology, using broad-scale datasets of water flow and water quality stations in catchments in France (Dupas *et al.*, 2015).

The first step, the PCA, requires normally distributed data so continuous catchment characteristics that were non-parametric (catchment area, circularity, mean elevation, standard deviation of elevation, TPI, TWI, drainage density and mean annual rain; see **Table 3.2**) were logarithmically transformed. All categorical variables were non-parametric due to the high numbers of zeros and were arc sine square-root transformed.

In HCPC, PCA is a pre-processing step where only principal components (PCs) with eigenvalues greater than one are carried through to the hierarchical clustering and partitional clustering steps (Cericola *et al.*, 2013) so that the signal of these data are maintained but the noise removed (Husson *et al.*, 2010).

A hierarchical clustering algorithm is then applied to the retained PCs to group similar clusters together. Euclidean distance is used to measure distance between classes and Ward's agglomeration method. The HCPC analysis is performed using the '*FactoMineR*' v1.41 package in R (Husson *et al.*, 2018).

The number of clusters from the hierarchical clustering is determined mathematically using the Davies–Bouldin index (as in the paper, **Figure 3.2**) but both in the paper and in other studies, expert interpretation of the output dendrogram is also needed for deciding the number of clusters (Azam *et al.*, 2017). This is followed by k-means partitioning clustering consolidation to create the selected number of clusters. The use of two clustering techniques increases the robustness of the final clusters reducing subjective tuning in cluster analysis (Azam *et al.*, 2017).

3.3.1.2 Interpreting HCPC outputs

The PCA output has eight PCs with eigenvalues greater than one, explaining 76% of the variance (**Figure 3.4a**). PC1 represents the gradient from upland to lowland and PC2 represents a gradient from highly coupled waterbodies dominated by hillslope processes to less coupled waterbodies. The gradients in this PCA are similar to the gradients observed in the SOM heatmaps in **Figure 3.1**.

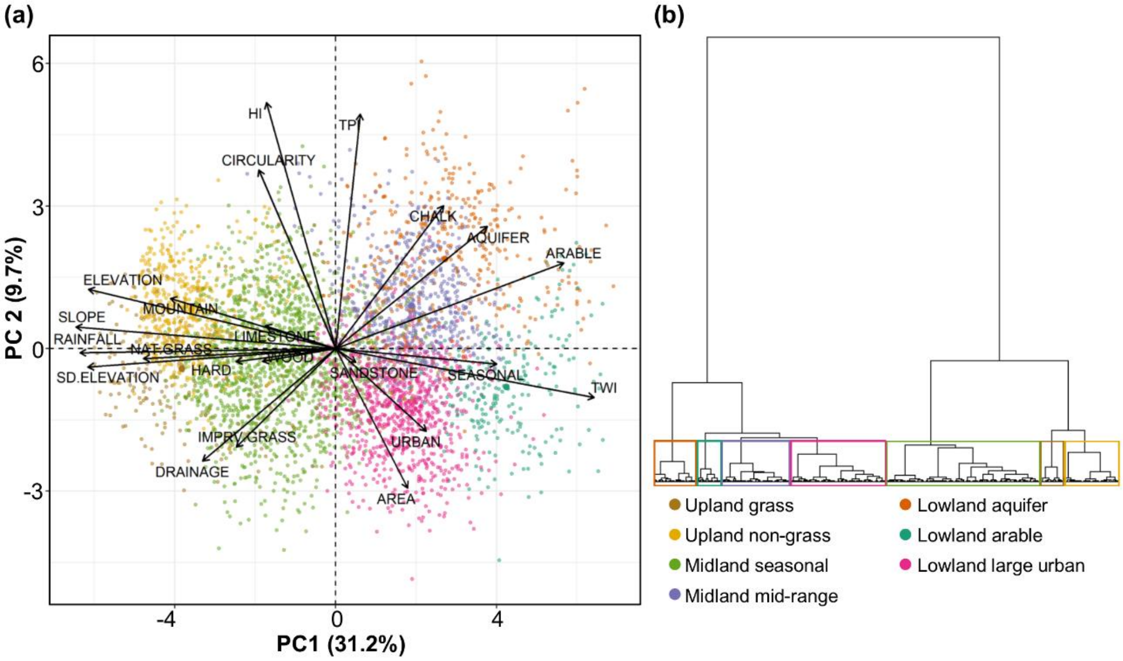


Figure 3.4. HCPC output: (a) Clusters following k-means consolidation plotted on factor map showing the PCA loadings of each catchment characteristic. (b) Hierarchical clustering dendrogram with seven clusters highlighted in coloured boxes.

A hierarchical clustering tree was built from the PCA output (**Figure 3.4b**). The characteristics of each cluster are determined by observation of the biplot (**Figure 3.4a**) and the median value of each characteristic within each cluster.

The optimum number of clusters for the PCA output was five according to the Davies-Bouldin index (**Figure 3.5**). However, if five clusters were selected here, the cluster types 'large urban', 'lowland arable' and 'mid-range catchment' would be combined into one class , or if six were selected 'mid-range' and 'large urban' would be combined (**Figure 3.4**). This is despite known different responses in rivers with different types and amounts of anthropogenic influences (Dadson *et al.*, 2017; Rothwell *et al.*, 2010). Seven types were therefore selected to separate these types and to be consistent with the SOM analysis for easy comparison.

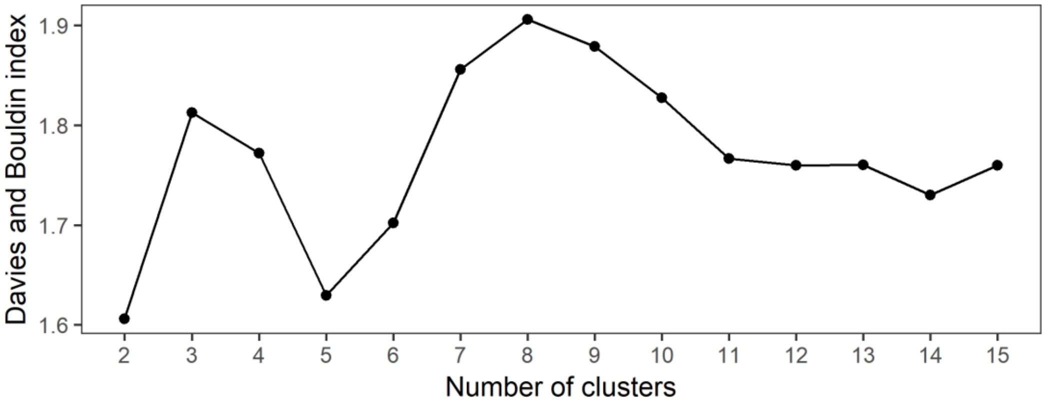


Figure 3.5. *Davies-Bouldin Index for the PCA output. The optimum number of clusters has the lowest index value. In this case five clusters are optimum, as two clusters is deemed too few to capture sufficient variation in catchment functioning.*

The seven clusters identified by the HCPC reflect seven waterbody types (**Figure 3.6a**) which are compared to the seven SOM-derived waterbody types (**Figure 3.6b**). The names attributed to each class reflect the catchment characteristics of the waterbody type. The characteristics are similar between the HCPC-derived and SOM-derived types, although there are differences highlighted in the maps in **Figure 3.6**.

The first division of the hierarchical tree is between upland and lowland waterbody types (**Figure 3.4b**). HCPC-derived types upland types – upland grass, upland non-grass and midland seasonal – have similar characteristics to SOM-derived types. Upland grass and upland non-grass have 386 and 293 fewer waterbodies in the HCPC-derived typology compared SOM-derived typology, respectively. In the HCPC typology the upland grass types are limited to igneous rock geology in the lake district and south west, and high relief areas of North Wales; and the upland non-grass waterbodies limited to the areas of highest relief. Whereas the SOM-derived upland waterbodies are more extensive across the higher elevation regions of England and Wales delineating known changes in elevation and geology, such as Exmoor, from surrounding lower elevation types (**Figure 3.6**). To compensate, the HCPC-derived typology has 1138 more waterbodies classified as midland seasonal that are extensively across high elevation regions, compared to the SOM-derived typology (**Figure 3.6**).

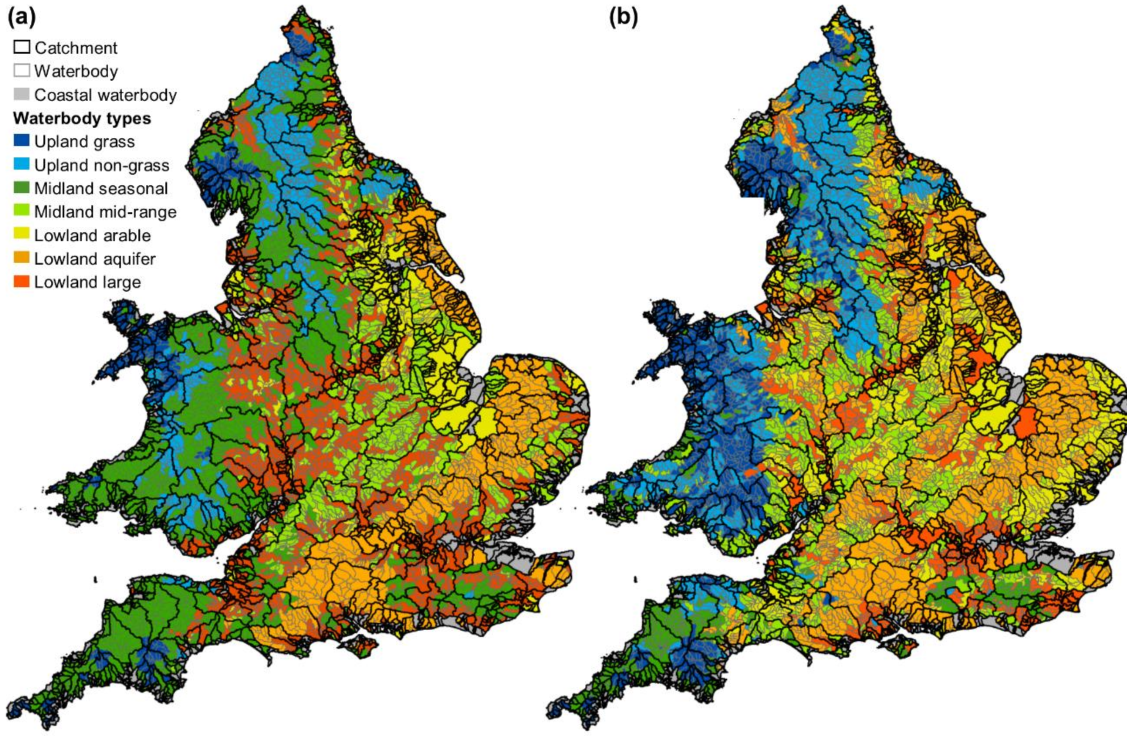


Figure 3.6. *Maps of waterbody typology for England and Wales based on the (a) HCPC and (b) SOM analysis. Black boundaries indicate catchment boundaries. The names attributed to each type is explained in the text.*

HCPC-derived lowland types – midland mid-range, lowland aquifer, large urban and lowland arable – have some differing characteristics and distributions compared to the SOM-derived typology. A similar number of waterbodies are classified as midland mid-range types by both methods. However, the HCPC-derived type follows bands of sandstone aquifer from south-west to north-east across England (**Figure 3.6a**), whereas the SOM-derived mid-range type is often adjacent to upland types or in central England (**Figure 3.6b**). This is because waterbodies with sandstone geology are classified in the lowland aquifer type by the SOM typology, whereas the HCPC-derived aquifer type is solely chalk geology and thus has 474 fewer waterbodies (**Figure 3.6**). The mid-range types in the SOM typology are classified as large urban by the HCPC typology and are more numerous extending across the lowlands (**Figure 3.6**). Many of the SOM-derived lowland arable types are also classified as large urban or mid-range by the HCPC-derived typology, restricting the range of lowland arable types to the lowest areas and sandstone geology of the east coast (**Figure 3.6**).

3.3.1.3 SOM as the superior method for this typology

The SOM technique is methodologically more appropriate for the creation of this typology of catchment controls than the HCPC method. SOMs have been compared to both PCA and hierarchical clustering as separate techniques for reducing complexity in large environmental datasets and, although the PCA and SOM analysis give comparable results, the SOM technique has additional advantages (Astel *et al.*, 2007).

The PCA component of the HCPC workflow required that input catchment characteristics have normal distribution and preferably not be co-correlated. This was problematic as most of the characteristics were non-normal and were inherently correlated, as is the case with many large environmental datasets (Feld *et al.*, 2016). The non-linear approach of SOM means the raw input characteristics may be used without transformation, retaining the topological structure of these data (ASCE, 2000). A previous study also praises the SOM technique for its ability to manage co-correlation between input environmental variables so as not to effect the clustering process (Bizzi and Lerner, 2012). However, the SOM is a black-box technique, so it is difficult to identify how the typology of the SOM map is discerned.

Both reduction techniques, PCA and SOM, reduce the dimensionality of these data into a two-dimensional visualisation, factorial maps (**Figure 3.4a**) or heatmaps (**Figure 3.1**) respectively. Analysing the heatmaps from the SOM provides a more holistic understanding of the characteristics of each type than when only considering the medians of each group and the distribution of the factorial map, as is the technique with HCPC. The heatmaps also enable the visualisation of both positive and negative correlations between variables and indicates outliers (Astel *et al.*, 2007).

Despite the differences in the data reduction technique, both methods produce waterbody types with similar defining characteristics that reflect a dominant upland-to-lowland gradient and a secondary heterogeneity gradient (**Figure 3.4a**). This reinforces the findings of the paper in Section 3.2.3, as similar gradients were identified by such differing methodologies. However, the number of waterbodies assigned to each type, and thus the resultant output maps in **Figure 3.6**, are very different. The SOM-derived typology is deemed superior as the maps in **Figure 3.6** better differentiate upland areas, urban centres and large rivers than the HCPC output. Also, geological features such as the sandstone aquifer are included in one class by the SOM typology but are split between two classes in the HCPC typology (although one would expect the functioning of catchments with sandstone aquifer geology to be similar). These differences stem from the ability of the SOM to account for the complex data structures present in the catchment characteristics by drawing type boundaries that are consistent with categorical boundaries and accounting for the non-transformed data values. It is therefore concluded that the SOM technique was the appropriate methodology for this task.

3.3.2 Application of the waterbody typology: Catchment complexity assessment

The typology developed in the paper in Section 3.2 is reflective of catchment-level effects on reach habitats. The distribution of the waterbody types indicates that there is variation in catchment-level effects, and by extension reach habitats, within catchment boundaries (**Figure 3.6b**). Waterbodies are part of a wider connected river network so catchment-level effects from upstream waterbodies will influence downstream waterbodies. However, waterbodies are classified in isolation in the typology presented in Section 3.2 and often waterbodies are considered in isolation of their surroundings by river managers where the focus is on improving the ecological and chemical quality of the individual waterbody. Therefore, there is an opportunity to use the waterbody typology developed in the published paper (Section 3.2) to explore the complexity of waterbody types within catchments, in terms of the diversity of waterbody types within the catchment and their connectivity.

Quantifying the spatial pattern of landscapes is common in the field of landscape ecology and the concept of 'riverscapes' embraces principles of landscape ecology to form a connected view of rivers (Fausch *et al.*, 2002; Wiens, 2002). In riverscapes, patches both within rivers, such as riffles and pools, and patches in the terrestrial catchment alter the movement of water, sediment, organisms and chemicals etc. in the riverine system (Wiens, 2002). Therefore, it is pertinent to quantify the spatial pattern of catchment characteristics within catchments.

Most landscape ecology studies are concerned with assessing the spatial pattern of patches in neighbourhoods. ***Patches*** are defined as discrete areas with relatively homogeneous properties (Kotliar and Wiens, 1990). ***Neighbourhoods*** are regions of interest within the landscape within which the spatial pattern of patches is quantified. In the present study the neighbourhood is the catchment boundary and patch properties are determined by waterbody type, so patches could be as small as a single waterbody or as large as an entire catchment.

An array of metrics have been used to quantify the heterogeneity of landscapes at both the patch and neighbourhood scales (Gustafson, 1998). While most frequently used as predictors of terrestrial species, the metrics have been applied in aquatic science to quantify land cover patterns and their influence on reach ecology, morphology and water chemistry (Hopkins, 2009; Wang *et al.*, 2001; Zhou *et al.*, 2012). However, other catchment characteristics are shown to influence the connectivity of the catchment, for example, steeper catchment topography has more effective sediment connectivity and erosion response than dissected or stepped landscapes (Baartman *et al.*, 2013).

In this brief application example, the spatial complexity of catchments across England and Wales is explored. Four example catchments (the Demonstration Test Catchments; more details in Chapter 4.2) are selected to understand how different spatial arrangements of catchment-level effects influence river reach features in more detail.

3.3.2.1 Delineating catchment boundaries

To explore within catchment complexity, catchment boundaries must be delineated for England and Wales. The typology is built on WFD waterbodies (see Section 3.2.2), sub-units of catchments, so catchment boundaries were created by merging waterbody polygons in ArcGIS based on the connectivity of the Centre for Ecology and Hydrology's (CEH) 1:50,000 blue-line network (Moore *et al.*, 1994). This technique was used as the existing WFD catchment boundaries, Surface Water Management Catchments (Environment Agency, 2014), encompassed multiple catchments that are physically unconnected by the river network (see **Figure 3.7** for an example). This arises because Management Catchments were designed to implement action plans for waterbodies rather than to represent the topographic catchment boundaries.

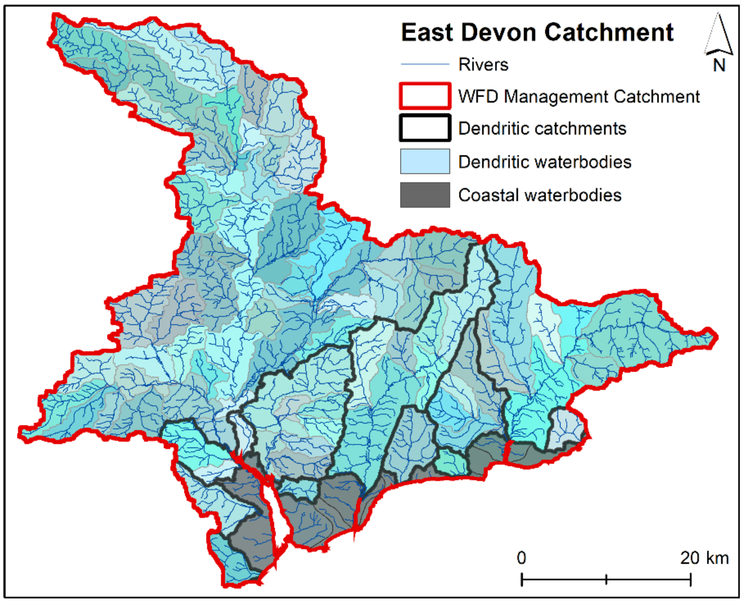


Figure 3.7. *The East Devon Surface Water Management Catchment. An example of how a Management Catchment may contain multiple dendritic catchments (as defined by this thesis) that are unconnected by the river network.*

Only certain waterbodies were retained to create the catchment boundaries used in this thesis, henceforth known as ***dendritic catchments***. Waterbodies that were defined as coastal by the WFD, large or modified (defined below) were not included in creating catchment boundaries (**Figure 3.8a**). Large and modified waterbodies are only excluded for the definition of catchment boundaries but are included in the waterbody typology.

Large waterbodies were defined as waterbodies with a major river. **Figure 3.8a** shows the large waterbodies in England and Wales, which track the paths of the major stems of the Tyne, Derwent, Ouse, Trent, Thames, Severn, Wye and Mersey. By removing waterbodies containing large rivers, large catchments may be split into their major sub-basins so that catchment-level effects were not generalised over too large an area. For example, the entire Thames River Basin takes up ~16% of the surface area of England but may be split into 23 dendritic catchments upstream of Windsor once large waterbodies are removed (**Figure 3.8c**).

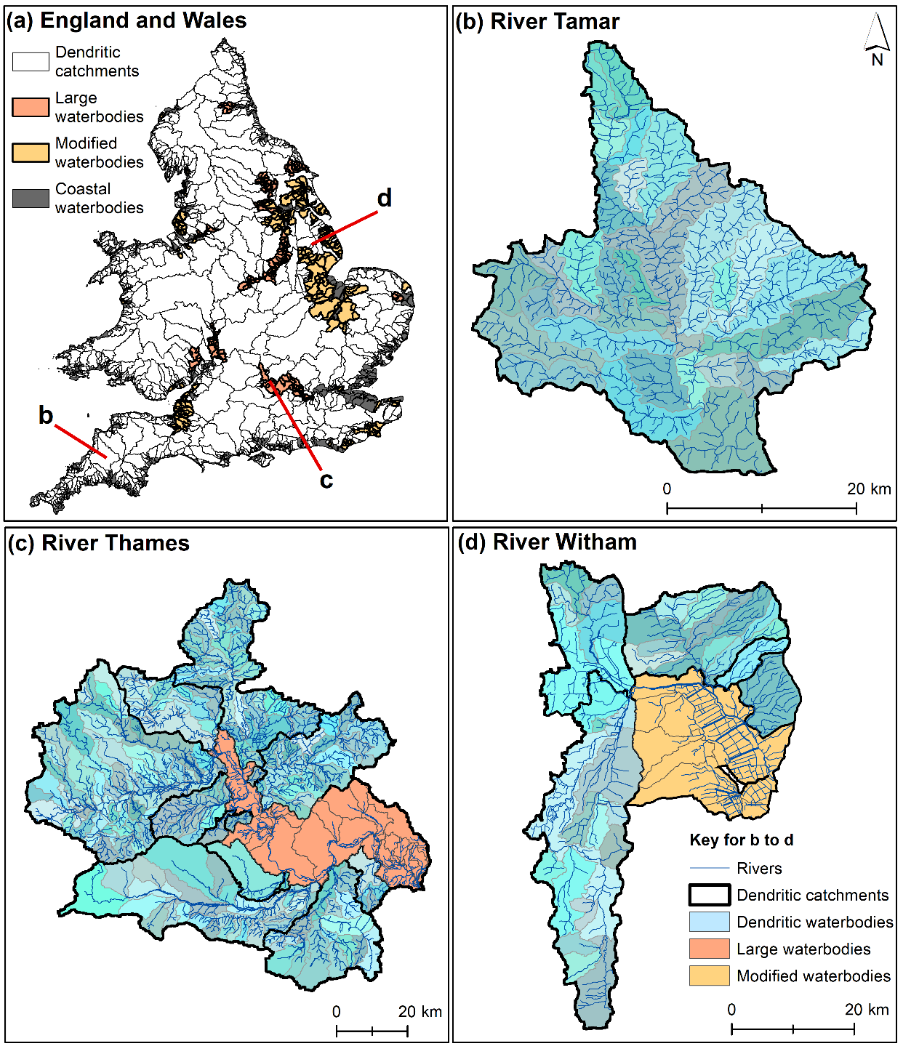


Figure 3.8. *Method for defining catchment boundaries: (a) Catchment boundaries in England and Wales. Waterbodies that are not in dendritic catchments indicated by colour. (b) Example catchment where all waterbodies are dendritic so are merged into one catchment. (c) Example catchment with large waterbodies. (d) Example catchment with modified waterbodies.*

Modified waterbodies were defined as waterbodies where the river network is no longer dendritic and has been modified beyond natural conditions. Modified waterbodies are recognised visually by identifying a grid-like network in the waterbody (see **Figure 3.8d** for example). The grid-like network is often indicative of land drainage, modifications in urban centres or water transfer schemes. For example, **Figure 3.8a** shows that these areas are concentrated in the east of England where there is considerable historic land drainage, and around the coastline. These waterbodies were removed because the artificial network connects catchments that may otherwise not be connected. Furthermore, the analysis on network structure conducted in Chapter 4 demands a dendritic network (see **Appendix 4A** for details), so removing modified catchments here ensures that analysis stays consistent throughout this thesis.

3.3.2.2 Quantifying catchment complexity nationally

The number of waterbody types within a catchment (*nTypes*) and the number of patches within a catchment (*nPatches*) were extracted in a GIS. A patch was defined as contiguous waterbodies of the same type. To gain a measure of catchment complexity a complexity index was calculated:

$$Complexity = 1 - \left(\frac{nTypes}{nPatches}\right)$$

(Equation 3.1)

High values indicate more complex catchments whereas low values indicate simpler catchments. Complexity was calculated for all 734 dendritic catchments. Of these catchments, 449 had only one waterbody and 524 catchments only had one waterbody type. Therefore, most catchments had a complexity of zero. The catchments that only had one patch are indicated in **Figure 3.9a**.

3.3.2.3 Patterns of catchment complexity within and between catchments

Catchment complexity is greater in larger catchments in central England and Wales (**Figure 3.9a**). This is likely because larger catchment exhibit the greater variation, draining a range of waterbody types (**Figure 3.9b**). Fröhlich *et al.* (2008) highlight that largest catchments can be the most heterogeneous in terms of land cover, geology etc. so become the most simplified when catchments are considered as single entities. For example, hydrological and chemical fluxes measured at the catchment outlet were integrative but not representative of the internal hydrological and chemical fluxes in the catchment. This highlights the dangers of considering catchments, particularly large catchments, as homogeneous areas when there is a range of internal variation.

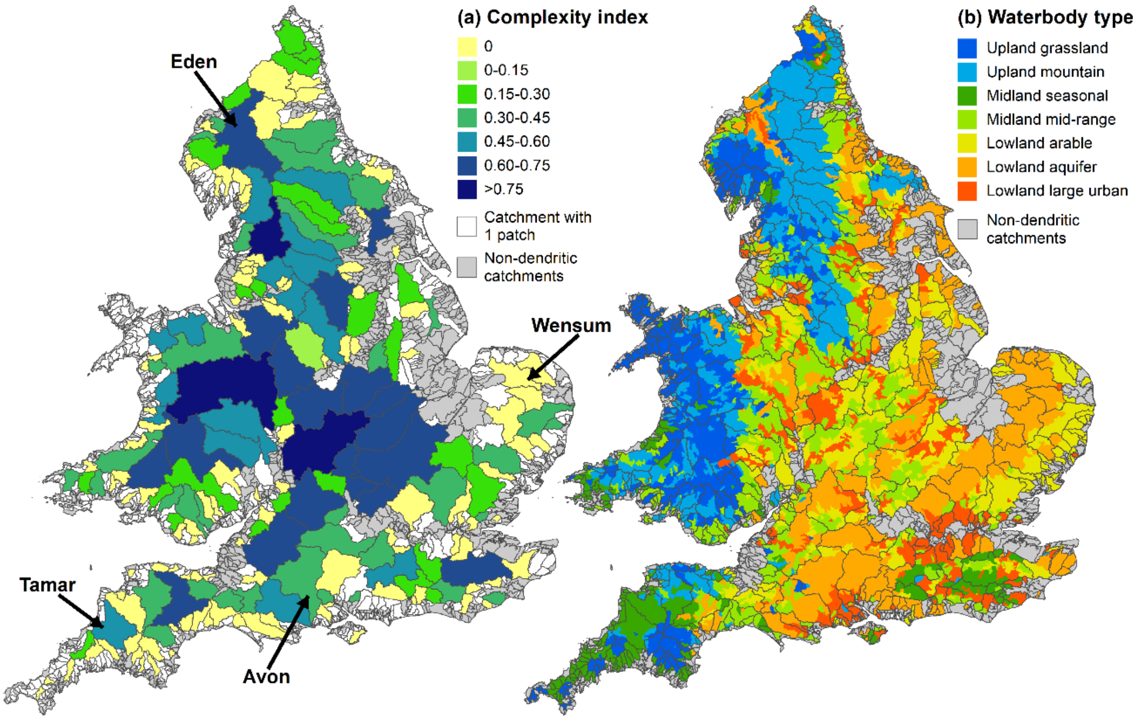


***Figure 3.9**. Catchment complexity mapped across England and Wales: (a) Complexity index (Equation 3.1) value distribution for catchments in England and Wales. Non-dendritic catchments (i.e. waterbodies that are coastal, large or modified; see Section 3.3.3.1) shown in grey and catchments containing one waterbody type shown in white. Demonstration Test Catchments labelled and explored in detail in **Figure 3.10**. (b) Typology of catchment-level effects for waterbodies in England and Wales for comparison (reprint of **Figure 3.2a** in paper with catchment boundaries).*

Catchments containing upland waterbodies tend to have relatively large patches and a greater number of waterbody types (**Figure 3.9**) because catchments with upland waterbodies can exhibit a full gradient from upland to lowland waterbody types. This means catchments with upland waterbodies are likely to have a wider range of processes, from highly coupled zones of sediment supply in upland sites to decoupled zones of sediment storage downstream (Schumm, 1977; **Figure 2.2**). In more lowland catchments, the full gradient from upland to lowland waterbody types cannot occur so there are often fewer waterbody types than in upland catchments. However, mid-range, lowland and large urban types are less continuous forming a greater number of discrete patches especially for areas in central England (**Figure 3.9a**). This supports observations of Turner and Ruscher (1988) that found lower landscape diversity in the uplands than the lowlands, however, the lowlands became less complex with larger patches as human influence increased, contrary to these results. Aquifer waterbodies occur in large patches due to the dominance of the geology variables in their classification. Catchments dominated by aquifer types, therefore, often have low complexity (**Figure 3.9**).

Upon visual inspection of the maps in **Figure 3.9**, there is a link between catchment complexity and catchment shape. Elongated catchments, such as those in the Eastern

Pennines, have linear bands of contiguous waterbody types from upstream to downstream. This may be because the restricted area limits the number of patches (Gardner *et al.*, 1987) as there are fewer waterbodies across the width of the catchment. However, catchments in central England are more circular with distinct sub-basins that may contain different catchment-level effects. It would be interesting to test this hypothesis that elongated upland catchments have more types but fewer patches and more circular lowland catchments have more patches but fewer types.

When exploring the four Demonstration Test Catchments in more detail (locations are shown in **Figure 3.9)** the national patterns that are observed emerge more clearly. The Wensum has zero complexity, being dominated by aquifer waterbodies that make up one patch, feeding into a large urban waterbody downstream. The Avon also has low complexity, despite containing six out of a possible seven waterbody types. This is because the waterbody types are contiguous forming larger patches, with little variation between sub-catchments.

The Tamar, on the other hand, is a circular catchment and despite exhibiting only three waterbody types, has a higher complexity score than the Avon. This is because it has distinct sub-catchments, with its headwaters in the upland region of Dartmoor to the east and in the upland region of Bodmin Moor to the west. This creates more dissected patches in a catchment that is dominated by one homogeneous patch of midland seasonal type. The Eden is the most complex, with sixteen different patches radiating from the main channel. It is also a relatively circular waterbody, exhibiting major sub-catchments that drain into the main aquifer patch (**Figure 3.10**).

Considering the complexity of specific catchments may aid river managers to assess the impacts of catchment-level effects not just in the waterbody of interest but for the upstream catchment. The complexity metric developed here is too broad for operational management of river reaches. Instead, it helps identify the types of catchment where an assessment of the range of catchment-level effects influencing a waterbody may be of use, such as large and/or circular catchments with multiple waterbody types. A simple assessment of the typology for individual catchments may be sufficient (e.g. **Figure 3.10**) however, a measure of connectivity between waterbodies would be beneficial. This is a challenge when quantifying the complexity of the catchment because waterbodies are not only connected across space but also directionally based on the hydraulic movement of water in the catchment in numerous dimensions (Petts and Amoros, 1996). Exploring directional connectivity of terrestrial landscape patches to the aquatic environment, particularly longitudinal and lateral connectivity would be an interesting development of this work and would aid understanding of the mechanisms connecting catchment-level effects to river reaches.

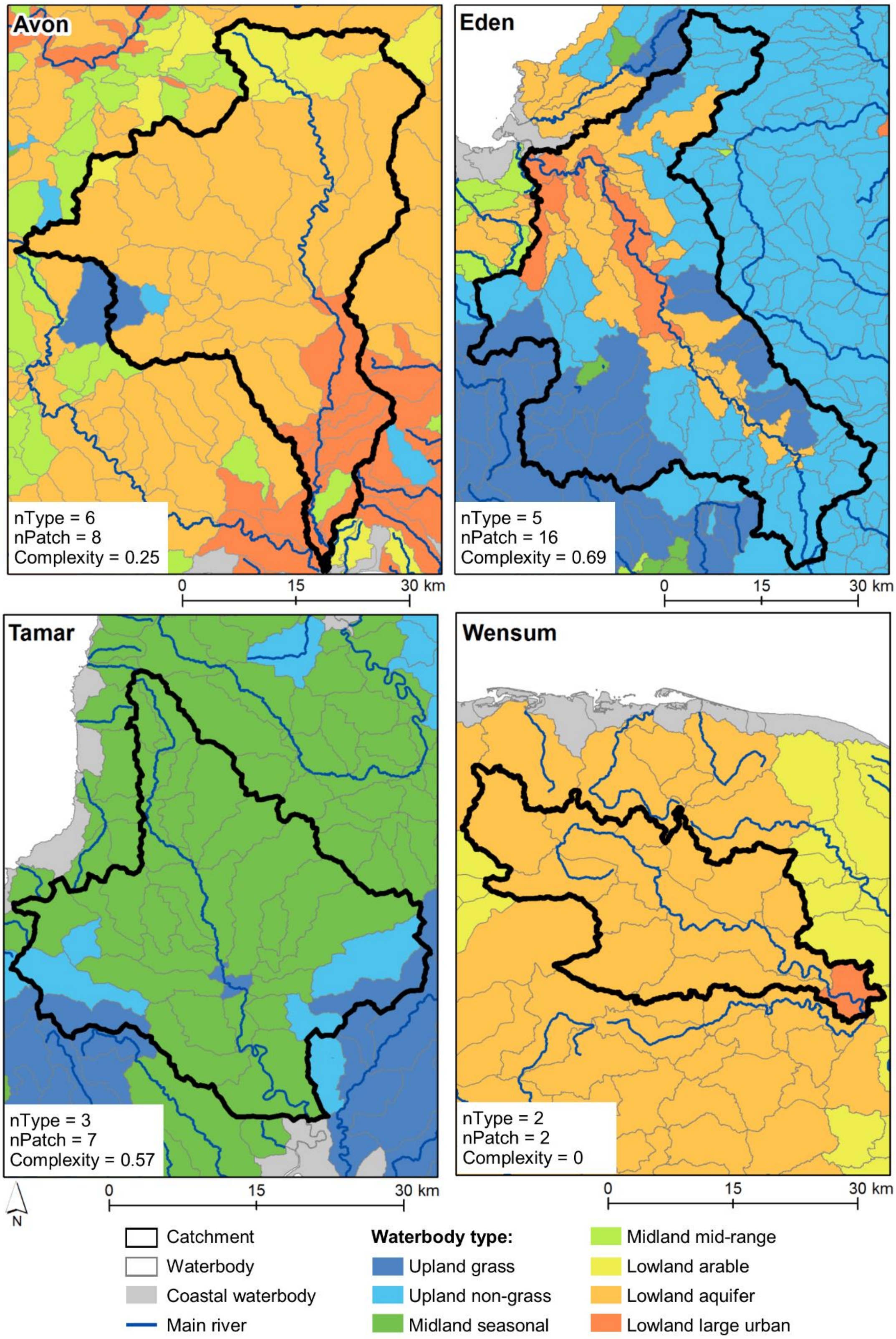


Figure 3.10. *Waterbody types in each Demonstration Test Catchment. Number of waterbody types (nType), the number of patches (nPatch) and the complexity index for each catchment are in the corner of each pane.*

3.4 CHAPTER CONCLUSIONS

Catchment-level effects can be classified into a useful number of waterbody types at a national level (**Figure 3.2**) that capture enough variation in multiple characteristics of the catchment to explain broad differences in physical habitats between types. This builds on the work of previous large-scale studies that have focused on anthropogenic catchment controls, particularly land cover (see evidence review in Section 2.1.3), to understand negative influences on river and aid restorative management of the system. Here, the machine learning technique, SOM, was implemented so that a range of natural and anthropogenic characteristics that control river reach functioning may be included in the typology (**Table 3.2**).

The SOM approach better accounted for the complexities of the catchment characteristic dataset than traditional methods of data reduction such as PCA (Section 3.3.2). The output heatmaps of the SOM visually aid the intuitive interpretation of correlations and outliers (**Figure 3.1**) in addition to the gradients in these data identified by PCA, making it a useful tool for typology creation and communication.

The typology created uses GIS-derived catchment controls to infer reach-level form and function which is contrary to most river typologies that classify the reach. This typology complements reach-level typologies as it quantifies control rather than response and is spatially continuous at a national level making it quick to update as new data becomes available. This makes the typology a useful tool for rapid assessment of catchment-level effects in a single waterbody, comparisons between waterbodies and for more strategic national planning by river managers where time and resources are often too limited to assess catchment-level effects in detail. The typology is applicable to river management as it utilises datasets and spatial levels that are relevant to regulatory compliance, for example, catchment-level effects are classified within WFD waterbodies and the typology is evaluated with physical habitat indices derived from the RHS. This may encourage wider appreciation of catchment-level effects by river managers working at the reach-level.

One application is explored here; identifying the complexity of waterbody types within the wider catchment (Section 3.3.3). It highlights the utility of looking beyond the waterbody unit, that is so often the unit of focus for managers, to the entire catchment network (**Figure 3.10**). Future research should explore the connectivity of waterbodies along the river network, by quantifying the combined effect of upstream waterbodies, which is necessary for a full appreciation of catchment-level effects.

CHAPTER 4

Integrating network topology metrics into studies of catchment-level effects on river characteristics

4.1 CHAPTER INTRODUCTION

This chapter was published in the journal Hydrology and Earth System Sciences in March 2019 and is included as the published manuscript in Section 4.2. The gaps in knowledge and thesis objectives addressed by the paper are briefly summarised below. For more detailed explanation, see Section 4.2.

The chapter explores an over-looked component of the catchment: the structure of the river network, known as network topology. Network topology has been included in few studies of catchment controls on river systems (see evidence review in Section 2.1.3) yet it is critical to catchment functioning, transmitting influences from distal reaches of the catchment downstream towards the outlet. Quantifying connectivity through the river network may aid understanding of catchment-level effects as previous studies that link catchment controls to reach responses do not focus on *how* this occurs as *"…pathways of influence may not be as easily detected"* (Allan, 2004, p.271). Network topology has before been quantified in the hydrological sciences for flood hydrograph estimation but has rarely been studied as a catchment control on river reaches for the purposes of river research across a range of disciplines. Therefore, the specific objective of this chapter and the paper is:

Objective 4a: To quantify network topology within catchments by creating a metric fit for multiple disciplinary use.

The objective is met by adapting two methods of network topology quantification from the field of hydrological estimation, to reflect the density of the river network. The adapted metrics are compared to a traditional method of network topology quantification by testing associations between each network metric and physical habitat indices derived from the RHS dataset (**Table 2.2**) in four example catchments in England. Additional analysis of habitat diversity indices that were not included in the published paper is presented in Section 4.3. Further details of the methodological approach taken to produce a dendritic network for the four catchments which is required to calculate network density is provided in **Appendix 4A**.

4.2 PUBLISHED PAPER

Text, tables and figures copied directly from Heasley et al. (2019). Numbering of sections and figures changed to coincide with the thesis. References are included in the full reference list for the thesis. The published paper is accessible online at https://doi.org/10.5194/hess-23-2305-2019.

4.2.0 Abstract

The spatial arrangement of the river network is a fundamental characteristic of the catchment, acting as a conduit between catchment-level effects and reach morphology and ecology. Yet river network structure is often simplified to reflect an upstream-to-downstream gradient of river characteristics, commonly represented by stream order. The aim of this study is to quantify network topological structure using two network density metrics – one that represents network density over distance and the other over elevation – that can easily be extracted from digital elevation models and so may be applied to any catchment across the globe. These metrics should better account for the multi-dimensional nature of the catchment than stream order and be functionally applicable across geomorphological, hydrological and ecological attributes of the catchment. The functional utility of the metrics is assessed by appropriating monitoring data collected for regulatory compliance to explore patterns of river characteristics in relation to network topology. This method is applied to four comparatively low energy, anthropogenically modified catchments in the UK using river characteristics derived from England's River Habitat Survey database. The patterns in river characteristics explained by network density metrics are compared to stream order as a standard measure of topology.

The results indicate that the network density metrics offer a richer and functionally more relevant description of network topology than stream order, highlighting differences in the density and spatial arrangement of each catchment's internal network structure. Correlations between the network density metrics and river characteristics show that habitat quality score consistently increases with network density in all catchments as hypothesized. For other measures of river character – modification score, flow-type speed and sediment size – there are varying responses in different catchments to the two network density metrics. There are few significant correlations between stream order and the river characteristics, highlighting the limitations of stream order in accounting for network topology. Overall, the results suggest that network density metrics are more powerful measures which conceptually and functionally provide an improved method of accounting for the impacts of network topology on the fluvial system.

4.2.1 Introduction

Rivers are integrators of many elements of their catchments (Dovers and Day, 1988). Consequently, integrated catchment management has long been seen as the gold standard for river management and has been adopted in catchments across the globe (Newson, 2009). Research linking patterns of river reach characteristics to catchment-level functioning is currently focused on characteristics of the terrestrial catchment such as land cover, geology and topography (e.g. Cohen *et al.*, 1998; Harvey *et al.*, 2008; Jusik *et al.*, 2015; Naura *et al.*, 2016; Richards *et al.*, 1996, 1997). Yet "hot-spots" of activity within catchments are identified based on the hydrological connectivity of the catchment (Newson, 2010), a characteristic that is often neglected by catchment-level studies. This missing component of the catchment is critical for true integrated catchment management as the impacts of key management features (e.g. water, channel, land, ecology and human activity) are transmitted throughout the river network (Downs *et al.*, 1991). By investigating the impacts of hydrological connectivity on river form and function, our understanding of catchment functioning can become more holistic and beneficial to catchment management.

Effective catchment management rests not only on improving scientific understanding of river form and function across multiple scales, but also on better integration between the key disciplines of catchment studies: geomorphology, hydrology and ecology. This type of interdisciplinary approach is critical for understanding complex multi-casual relationships in river systems (Dollar *et al.*, 2007). However, catchment connectivity is parameterized differently by different disciplines based on their interests. The discipline of geomorphology focuses on characterizing the morphometry of the catchment, either using general variables which are continuous across the landscape (e.g. elevation, slope, curvature) or specific variables which represent individual features such as catchments (e.g. drainage density, shape, area) or streams (e.g. stream order, stream length) (Evans and Minár, 2011). Hydrology focuses on how the catchment influences hydrograph and flood peak timing and magnitude. Methods such as the Geomorphic Instantaneous Unit Hydrograph (Rodriguez-Iturbe and Valdes, 1979) focus on predicting the travel time of water reaching channels and travelling downstream based the morphology of the catchment, drainage network and precipitation. Aquatic ecology takes a network-centric approach, utilizing dendritic ecological networks (Peterson *et al.*, 2013). This method aims to take a spatially continuous view of rivers (Fausch *et al.*, 2002) in order to appreciate the influence of flow and location in the network on discrete sites chosen for ecological sampling. Spatial statistical stream network models based on the branching of the network

(Ver Hoef and Peterson, 2010) are shown to be more accurate than a standard Euclidean distance kriging model, yet only worthwhile if data sites are distributed across the network and are spatially correlated (Peterson *et al.*, 2013). Alternate methods for exploring relationships between network structure and ecological functioning are also based on Euclidean distance along the network (Ver Hoef and Peterson, 2010).

Each discipline represents the elements of the catchment critical to their field, focusing on describing catchment form, catchment flow responses and ecological responses. However, the geomorphology, hydrology and ecology of the catchment are interconnected across spatial and temporal dimensions in the fluvial hydrosystem (Petts and Amoros, 1996). We argue that the overlap between disciplinary methods can be utilized to create a metric to represent the catchment that is meaningful across all disciplines and offers increased potential for effective catchment management utilizing a multi- or inter-disciplinary approach.

This paper repurposes metrics that focus on the topology of the river network for a novel application: to assess the key link between the catchment and reach-level functioning. The metrics represent network density variation within catchments and have functional applications across the fields of geomorphology, hydrology and ecology (Section 4.2.1.1). The impacts of internal network structure on patterns of river characteristics within catchments are explored by utilizing datasets that are collected for regulatory purposes, with areas of higher network density likely to support greater river quality and diversity (Section 4.2.1.2). The utility of the topological metrics is compared against stream order, a classic but oversimplified method of accounting for network topology. The topology metrics are calculated for catchments with comparatively low energy and that are influenced by anthropogenic modification as much of the previous evidence for increases in diversity in network-dense areas has been from highly erosive mountainous catchments.

4.2.1.1 Quantifying the river network at different scales and dimensions

River network structure, or network topology, is one way to conceptualize the integrated transport of water, sediment and nutrients from the upstream catchment to downstream reaches. The spatial arrangement of links (river channels) and nodes (confluences) concentrates the catchment effect in some areas of the landscape, making network topology a useful archetype of catchment functioning (Gupta and Mesa, 1988).

Drainage density (the total length of the network divided by catchment area) is most commonly used to compare the amount of the catchment covered by river channels, but this fails to quantify spatial variation within catchments, and so offers only a partial means for functionally assessing catchment similarities and differences. To represent within catchment network structure stream order (Strahler, 1957), ordering river links along an upstream-to-downstream gradient based on their upstream connectivity (**Figure 4.1a**) is also commonly used. However, stream order does not account for the spatial arrangement of links, only their relative position in the distance dimension of the catchment. Conceptualizing the catchment in this one dimension leads also to oversimplification; for example, first-order streams are thought of as upland headwater streams, furthest away from the river mouth, yet often first-order streams are tributaries to high-order, lowland streams with different characteristics than upland streams.

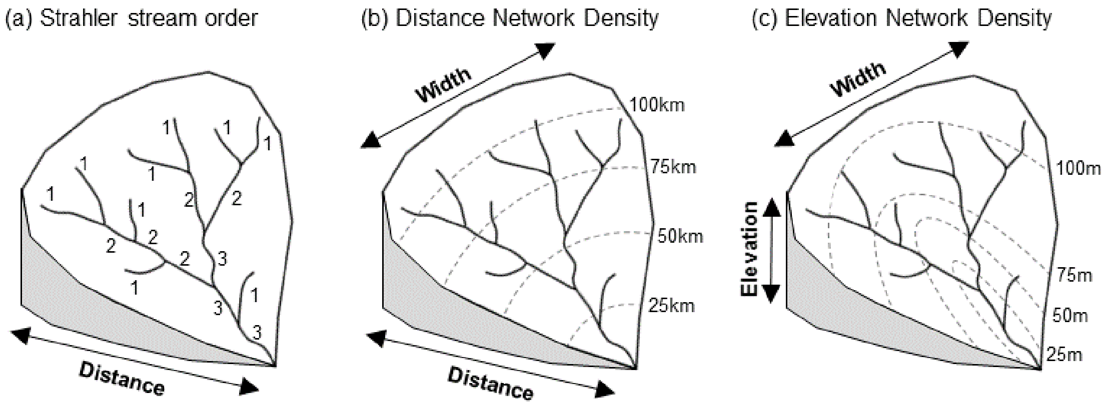


Figure 4.1. *Topological metrics explored in this paper and the dimensions of the network they represent. (a) Strahler stream ordering representing only the distance dimension of the network. (b) Distance network density representing the width dimension of the network at each distance interval (inspired by the network width function; Kirkby, 1976). (c) Elevation network density representing the width dimension of the network at each elevation interval (inspired by the link concentration function; Gupta et al., 1986).*

This paper argues that the spatial arrangement of links within catchments must be considered across the distance and height of the catchment to obtain a full three-dimensional appreciation of catchment effect through network topology. Two methods from the field of hydrology – network width function (NWF; Kirkby, 1976) and link concentration function (LCF; Gupta *et al.*, 1986) – offer increased dimensionality by accounting for the width of the network (i.e. the number of links) at successive distances, for the NWF or elevations, for the LCF.

These methods quantify network topology within catchments with functional significance. NWF has hydrological application, representing the travel time of water through the network to predict the timing and magnitude of unit hydrographs and flood peaks (Rodriguez-Iturbe and Valdes, 1979) with a more functionally specific method than the traditional stream-ordering approach (Gupta and Waymire, 1983). Extending applications beyond the field of hydrology, the timing and magnitude of the hydrograph have direct influence on instream ecology, controlling the formation, maintenance and disturbance of physical habitats (Bunn and Arthington, 2002). Longitudinal connectivity of water and sediment through the network is also one of the multiple dimensions of the fluvial hydrosystems approach to catchment ecohydrology (Petts and Amoros, 1996), influencing the capacity for lateral and vertical connectivity and the development of the riparian corridor over time. LCF is less frequently applied in hydrograph prediction than NWF. However, it may better reflect catchment hydrology by incorporating the effect of gradient on the travel time of water, rather than the constant travel time suggested by NWF (Gupta *et al.*, 1986). These metrics also have morphometric significance, reflecting the internal shape of the network by segmenting catchments into intervals to represent how network density changes within catchments (Stepinski and Stepinski, 2005).

This paper repurposes these metrics to reflect network density as a feature of the catchment rather than as a method for hydrograph prediction. Distance network density (modelled on the NWF) (**Figure 4.1b**) and elevation network density (modelled on the LCF) (**Figure 4.1c**) allow for the comparison and quantification of network topological variation both within and between catchment with improved interdisciplinary and functional applicability than the stream-ordering approach.

4.2.1.2 Network topology effects on river reach functioning

The topological structure of the river network configures the river ecosystem (Bravard and Gilvear, 1996) by impacting functioning at the reach and sub-reach scales. The distance dimension of the catchment, often represented by stream order (**Figure 4.1a**), reflects upstream-to-downstream gradual changes exhibited by many in-channel features and species. It forms the basis of classic geomorphic models, highlighting the zones of sediment supply in the headwaters, sediment transfer in the mid reaches and sediment storage near the outlet (Schumm, 1977). It is also a key component in classic ecological models such as the River Continuum Concept (Vannote *et al.*, 1980) which describes gradual changes in grain size, channel width, invertebrates, fish species and energy sources along the gradient. Both models suggest that diversity in channel morphology and biota may be highest in the mid reaches as channels transition from erosional to depositional environments. The River Continuum Concept is a popular model, but is critiqued for being too simplistic and for neglecting discontinuity introduced by changes at confluences (Perry and Schaeffer, 1987; Rice *et al.*, 2001). Confluences, as nodes in the network, are associated with changes in hydrological, geomorphological (Best, 1987; Church and

Kellerhals, 1978) and ecological (Kiffney *et al.*, 2006; Rice *et al.*, 2001) conditions and have therefore been termed biodiversity "hotspots" (Benda *et al.*, 2004b).

Confluence impacts extend throughout the river network, with increased channel heterogeneity in the tributary and main channel upstream and downstream of the confluence (Rice, 2017). This has led to several theories relating to the impact of numerous confluences in the context of the wider network. The Link Discontinuity Concept shows the impact of confluences throughout the length on the main channel, creating step changes in sediment size before fining continues downstream towards the next confluence along a "sedimentary link" (Rice *et al.*, 2001). The Network Dynamics Hypothesis posits that catchments with higher drainage density, and thus more confluences, will have greater channel heterogeneity (Benda *et al.*, 2004b), despite drainage density failing to be a useful catchment characteristic for predicting local habitat features (Davies *et al.*, 2000), 2000). The hypothesis also suggests that catchment shape will influence the impact of confluences, as more compact catchments will have more similarly sized tributaries (Benda *et al.*, 2004b), which have the greatest impact on channel morphology (Benda *et al.*, 2004a), the greatest flow diversity (Schindfessel *et al.*, 2015), and the greatest fish community diversity (Osborne and Wiley, 1992). In contrast, others have found that tributaries that differ most in size have the greatest impact. For example, Jones and Schmidt (2016) suggest that high densities of small tributaries flowing into a large channel cause small, cumulative changes, and Milesi and Melo's (2013) study concluded that small tributaries flowing into large channels in the peripheral regions of the catchment have the greatest impact on macroinvertebrate assemblages.

Interestingly, there is little evidence of anthropogenic impacts at confluences in the literature, but as confluences are proposed concentration points of catchment effects, it seems likely that they may be focal points for anthropogenic impacts. For example, flood events may occur downstream of large confluences as flood peaks converge, creating the need for flood defence measures (Depettris *et al.*, 2000), and scour and erosion at confluence junctions (Best, 1986) increase the need for bed and bank protection. Also, sediment size at confluences is shown to increase in many studies (Church and Kellerhals, 1978; Knighton, 1980), but in tributaries whose watersheds are dominated by agricultural land uses, fine sediments may become dominant at confluences, potentially altering river functioning (Owens *et al.*, 2005).

Many previous studies citing the impact of the network, specifically confluences, on river characteristics were conducted in highly erosive, relatively natural environments

(Network Dynamics Hypothesis, Benda *et al.*, 2004b; Link Discontinuity Concept, Rice *et al.*, 2001). Therefore, it will be interesting to assess the impact of network structure on river characteristics in catchments in England, a landscape that has undergone modification impacting catchment functioning for centuries (Macklin and Lewin, 2003).

4.2.2 Methods

4.2.2.1 Study sites

Four catchments are selected for testing the impact of network topology on river characteristics in England. The catchments are from the Demonstration Test Catchment programme (**Figure 4.2**) which are representative of 80% of soil and rainfall combinations in the UK (McGonigle *et al.*, 2014). This demonstrates the potential use of topological metrics for catchments with varying geologies and land uses. The Avon and Wensum catchments have similar characteristics, both being dominated by chalk geology with lower average annual rainfall and a high percentage of arable farming land cover. In comparison, the Eden and Tamar are dominated by less permeable bedrock with higher average annual rainfall and a high percentage of grassland land covers. In terms of their morphometry, the Avon and Wensum both have an elongated shape and low drainage density. The Wensum has the lowest relief with a maximum elevation of 95 m. The Tamar has the smallest catchment area (928 km^2) and is the most circular. The Eden is the largest catchment (2295 km^2) and has the highest maximum elevation (246 m).

4.2.2.2 Network topology metrics

Network topology metrics were calculated for each catchment using the 1:50,000 river network map, derived from both a digital terrain model (DTM) and Ordnance Survey data (Moore *et al.*, 1994). Anabranches and incorrectly digitized links in the network are identified using RivEX (Hornby, 2010) and removed. Removing anabranches was necessary as the topological metrics were designed for dendritic networks, so multi-thread channels, either naturally occurring or artificial ditches, would distort the calculations. This resulted in a total of 448, 2812, 1516 and 532 links in the Avon, Eden, Tamar and Wensum, respectively.

Elevation data were extracted from the Integrated Hydrological DTM (Morris and Flavin, 1994), a 50x50m gridded elevation raster with a 10 cm vertical resolution. Average elevation of each link and the distance from each link to the network outlet were extracted using RivEX (Hornby, 2010).

To extract a measure of network density that varies spatially within the catchment, each network is divided into 20 intervals, each of which represents 5% of the total distance or highest elevation in the network (**Figure 4.2**).

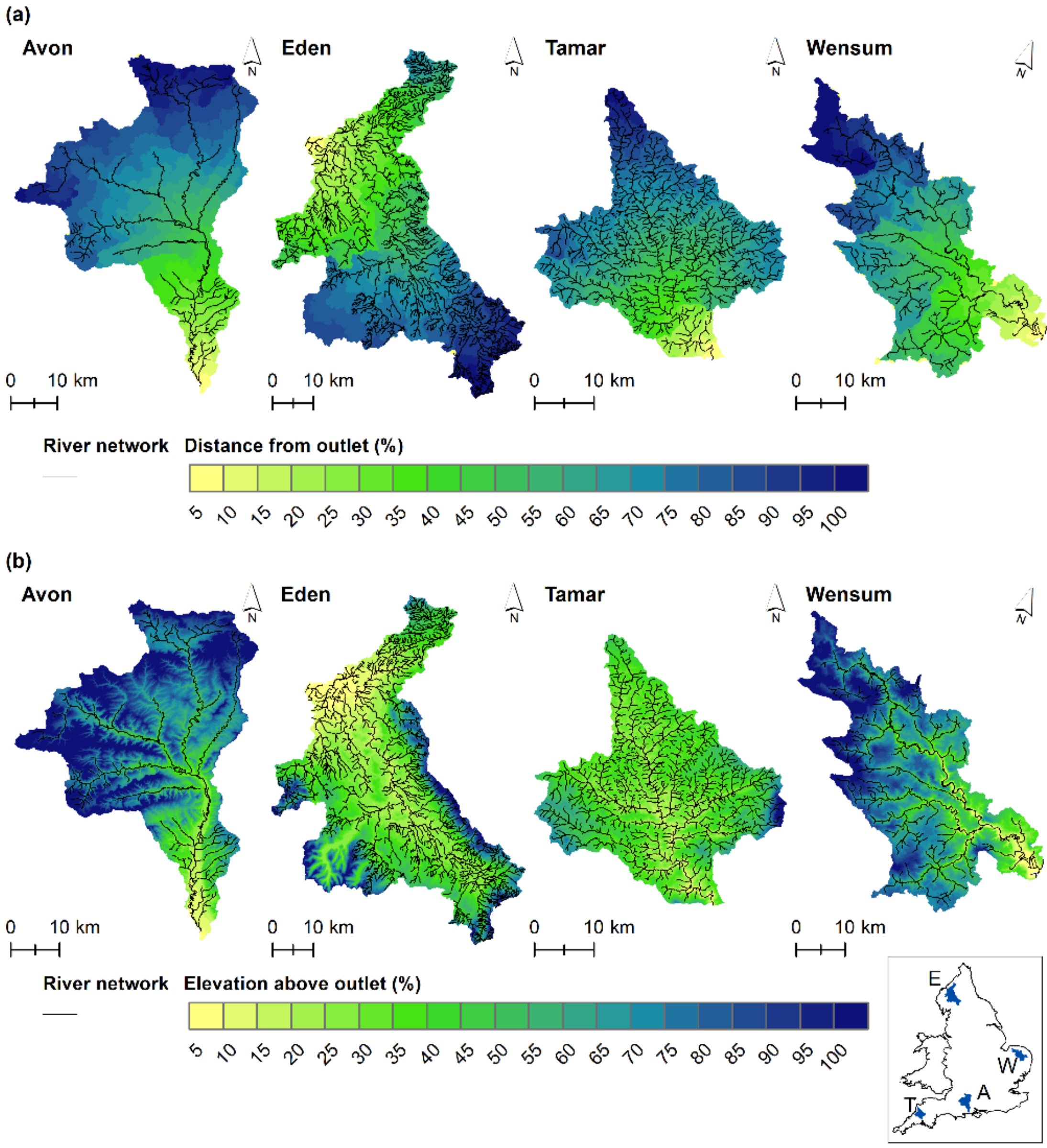


Figure 4.2. *Distance and elevation intervals for each Demonstration Test Catchment: Avon (A), Eden (E), Tamar (T) and Wensum (W). (a) Percentage distance intervals used to calculate distance network density. (b) Percentage elevation intervals used to calculate elevation network density. Map of catchment locations in England in the bottom-right corner.*

The network is divided in this manner based on the methods of the NWF and LCF, which have functional application to hydrograph prediction. Twenty intervals provide a relatively coarse sampling of the network, compared to the 100 intervals described by Stepinski and Stepinski (2005) when they adapted a morphometric variable, circularity ratio, to represent internal catchment elongation. Here, a total of 20 intervals is chosen so that most intervals contain links for the density calculation whilst ensuring the spatial distribution of network density within the catchment is characterized. Distance network density was calculated following

$$Distance\ network\ density = \frac{[n(d_0), \dots, n(d_i), \dots, n(d_N)]}{(d_N \times 0.05)}$$

(Equation 4.1)

where the number of links (n( )) within each 5% distance interval (d_i) from the outlet (d_o) to the maximum distance in the network (d_N) is normalized by the width of the interval (d_N x 0.05).

Elevation network density was calculated following

$$Elevation\ network\ density = \frac{[n(z_0), \dots, n(z_i), \dots, n(z_N)]}{(z_N \times 0.05)}$$

(Equation 4.2)

where the number of links (n( )) within each 5% elevation interval (z_i) from the outlet (z_o) to the maximum height of the network (z_N) is normalized by the width of the interval (z_N x 0.05). Normalization allows network densities to be compared between catchments controlling for differences in size and elevation as well as within catchments.

To assess the utility of the multi-dimensional topology metrics in accounting for the spatial structure of the network, the metrics are compared to the one-dimensional Strahler stream-order metric, extracted from the river network dataset using RivEX (Hornby, 2010).

4.2.2.3 River characteristics

The impact of network topology on channel functioning is explored using a broad-scale approach, i.e. adapting data collected for regulatory compliance to answer scientific questions. Adapting such datasets to scientific enquiry allows analysis to be conducted in many catchments across a wide spatial extent. There are many habitat monitoring methods across the globe, with 121 survey methods recorded in over 26 different countries (Belletti *et al.*, 2015), so this method may be adapted to other countries.

This study utilizes the River Habitat Survey (RHS; Raven *et al.*, 1996), a regulatory dataset collected by England's Environment Agency, which is used to reflect the river reach characteristics in each catchment. This dataset has been used to identify catchment effects on river characteristics in broadscale studies by previous research (e.g. Harvey *et al.*, 2008; Naura *et al.*, 2016; Vaughan *et al.*, 2013), but none have included the effects of network topology.

Since 1994, over 24,000 sites have been sampled in catchments across England and Wales, including the Avon (n=418), Eden (n=398), Tamar (n=189) and Wensum (n=315). Surveys were conducted at random sites within each 10 km² of England and Wales to ensure geographic coverage; however, this produces sampling bias as streams in high-density areas will be under-represented in the dataset, which is acknowledged in this study and discussed below.

At each site, over 100 features are recorded along a 500m reach with 10 "spot-check" surveys conducted every 50m and a "sweep-up" survey conducted across the whole reach (see Raven *et al.*, 1996, for details). Variables of interest that are hypothesized to be impacted by network structure can be calculated from the RHS observations (**Table 4.1**).

Table 4.1. *RHS variables calculated from RHS observations.*

RHS Variable	Calculation from RHS observations	Units
Habitat quality Assessment (HQA)	A score indicating the degree of naturalness and diversity of the riparian zone based on observations in the reach of flow types, substrate, channel and bank features, riparian vegetation etc.	HQA scale
Habitat Modification Score (HMS)	A score indicating the degree of artificial modification of the channel based on observations in the reach of reinforcements, re-sectioning, embankments, weed-cutting, realignment, culverts, dams, weirs etc.	HMS scale
Sediment size	$= \frac{(-8 \cdot BO - 7 \cdot CO - 3.5 \cdot GP - 1.5 \cdot SA + 1.5 \cdot SI + 9 \cdot CL)}{(BO + CO + GP + SA + SI + CL)}$ BO (boulder), CO (cobble), GP (gravel-pebble), SA (sand), SI (silt) and CL (clay) represent the number of spot checks allocated to each sediment size class	Approx. phi units
Flow type speed	$= \frac{(0 \cdot DR + 1 \cdot NP + 2 \cdot UP + 3 \cdot SM + 4 \cdot RP + 5 \cdot UW + 6 \cdot BW + 7 \cdot CF + 8 \cdot CH + 9 \cdot FF)}{(DR + NP + UP + SM + RP + UW + BW + CF + CH + FF)}$ DR (dry), NP (no perceptible flow), UP (upwelling), SM (smooth), RP (rippled), UW (unbroken wave), BW (broken wave), CF (chaotic flow), CH (chute), FF (free-fall) represent the number of spot checks allocated to each flow speed class	Flow type speed scale

The Habitat Quality Assessment (HQA) and Habitat Modification Score (HMS) variables are both amalgamations of RHS observations with individual features given a score derived by expert opinion (see Raven *et al.*, 1998, for more details). The scoring systems are subjective, but HQA and HMS provide overviews of the channel condition that are widely applied for regulatory compliance. The scores are therefore included in this study to reflect how they may be impacted by network topology.

The remaining RHS variables are calculated directly from RHS observations and so are more objective. Sediment size is calculated as a reach average of spot-check observations using the same method as previous studies (Davenport *et al.*, 2004; Emery *et al.*, 2004;

Harvey *et al.*, 2008b). Flow-type speed was calculated in the same manner as sediment size using values of flow which represent an approximate flow velocity gradient defined in Davenport *et al.* (2004). These variables were chosen to reflect dominant geomorphic processes occurring in each reach and due to the prominence of sediment size and flow type in defining physical habitats for instream biota (Rowntree and Wadeson, 1996). The variables are likely to be impacted by the density of the river network as they have been shown to be impacted by individual confluences. For example, channels are shown to become more geomorphologically heterogeneous (Benda *et al.*, 2004a) and s size has been shown to coarsen at confluences (Rice *et al.*, 2001). Surface flow types are also likely to become more diverse at confluences as the convergence of channels creates a number of different flow environments (Best, 1985) that result in different water-surface topographies (Biron *et al.*, 2002).

It must be noted that the RHS dataset was collected for regulatory compliance and was not directly intended for scientific enquiry. Therefore, there is a limitation in the amount of detail that can be extracted about physical processes as the observations recorded are an average across a 500m reach. Despite this, the wide spatial coverage of the dataset makes it a powerful tool, allowing analysis to be conducted across multiple catchments with differing characteristics.

For each distance and elevation interval created by the network topology metrics, descriptive statistics (mean, median,90th and 10th percentiles) of each RHS variable are calculated. Despite the RHS sampling strategy (Jeffers, 1998b) biasing site selection towards less dense areas of the network, most distance and elevation intervals contained RHS sites (with only some low-density intervals not containing RHS sites). This method is designed to account for natural variation and modification at individual RHS sites, in order to assess broad patterns of reach characteristics at the catchment level.

4.2.2.4 Statistical analysis

Analysis is conducted with all catchments combined into a single population to identify overall trends across all catchments, a method used in previous broad-scale studies. The analysis is also split into individual catchments to identify how the relationship between network topology metrics and river reach characteristics differed between catchments.

Correlation tests are used to determine the strength and direction of the association between the descriptive statistics of the RHS variables and distance network density, elevation network density and stream order to ascertain how reach characteristics respond to network topology. Kendall's correlation method was used as the variables have non-normal distributions, a small sample size and tied data values (Helsel and Hirsch, 2002). The effect size of Kendall's τ is lower than other correlation methods with strong correlations occurring with τ values greater than 0.7 (Helsel and Hirsch, 2002).

As multiple correlations are conducted, false discovery rate (Benjamini and Hochberg, 1995) corrections were applied to the p-values produced from the Kendall correlations to reduce the risk of type I error. The false discovery rate method has been found to be more powerful than other procedures for controlling for multiple tests (Glickman *et al.*, 2014).

4.2.3 Results

4.2.3.1 Differences in network topology metrics between catchments

The topological metrics developed in this study show the internal structure of the network for each catchment. The separation of the catchments into distance and elevation intervals emphasizes different features of the catchment. The distance intervals (**Figure 4.2a**) are arranged longitudinally within the catchment, highlighting sub-basins within each catchment. The elevation intervals (**Figure 4.2b**) have a radial arrangement, centering around the incised main channel of each catchment.

Distance network density is higher in the Eden (28.4±10.3) and Tamar (44.1±21.9) compared to the Avon (4.7±1.9) and Wensum (6.8±0.7), interesting as the Tamar is the smallest catchment by area. The shape of the distance network density function reflects the internal shape of the network (**Figure 4.3a**). For example, the Tamar has a peaked density distribution reflecting the circular shape of the catchment such that the majority of links are at 55%–65% distance from the catchment outlet. The Avon and Eden reflect similar internal network structures, both exhibiting a bimodal density distribution, despite the differences in the number of links in the catchments. The density distribution of the Wensum has a more complex internal distribution of links with multiple peaks in density.

Elevation network density (**Figure 4.3b**) shows similar density distribution shapes to distance network density, with a unimodal distribution for the Tamar and multi-modal distributions in the other catchments. In contrast to distance network density, elevation network density shows the highest peaks in density in the Tamar (10.3±5.0) and Wensum (10.1±4.3), despite the Wensum having the lowest network elevation, and has lower values in the Avon (3.4±1.0) and Eden (5.8±2.5). The peak densities in the Wensum occur at similar positions in the elevation and distance intervals, whereas the peaks in the other catchments are negatively skewed, showing the network density is highest at low–mid elevations.

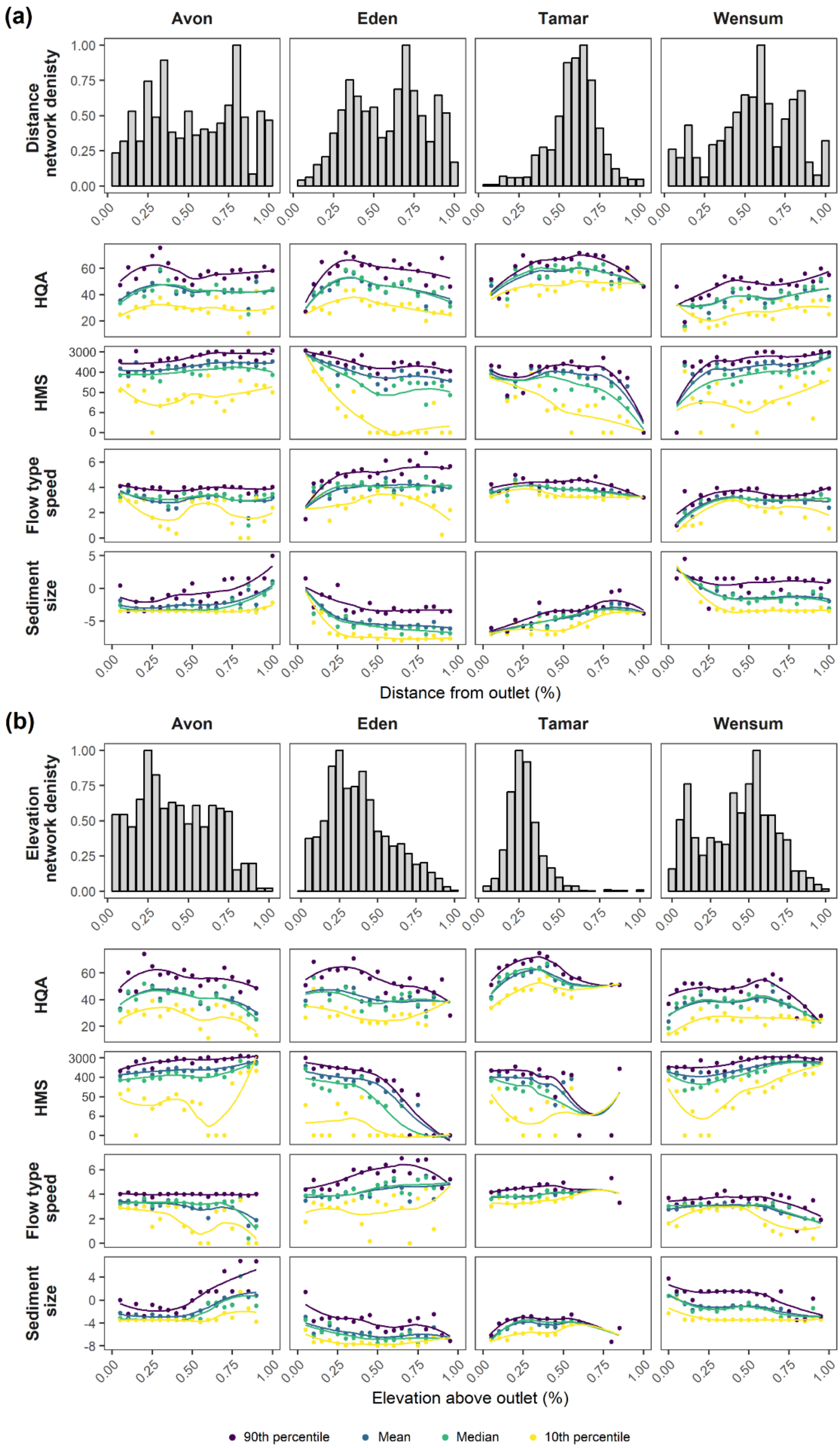


Figure 4.3. *Network topology metrics (a) distance network density and (b) elevation network density. Descriptive statistics of each RHS variable over (a) distance and (b) elevation for each catchment with smooth loess lines to indicate trend. Network topology metrics are normalized between 0 and 1 and HMS score is logarithmically transformed for display purposes.*

Nearly half of the links in each catchment are classified as first-order streams and the number of links declines exponentially towards the highest orders in all four catchments. There are weak correlations (τ=-0.03 to 0.17) between the three network topology metrics; distance network density, elevation network density and stream order. This suggests that the metrics are independent and reflect different aspects of river network topology.

4.2.3.2 *River characteristic relationships with network topology metrics*

RHS sites in the Avon and Wensum have similar river characteristics. Both have lower habitat quality, high modification, fine sediment and slower flow types than the Eden and Tamar. When all catchments are combined, there are significant ($p<0.05$ after p-value correction) correlations with most descriptive statistics for each RHS variable and distance network density (**Figure 4.4**). There are consistently positive correlations with HQA and flow-type speed and negative correlations with HMS and sediment size. There are fewer and weaker significant correlations with elevation network density (**Figure 4.4**). The only significant correlations with stream order are with HMS, which shows a negative correlation (**Figure 4.4**).

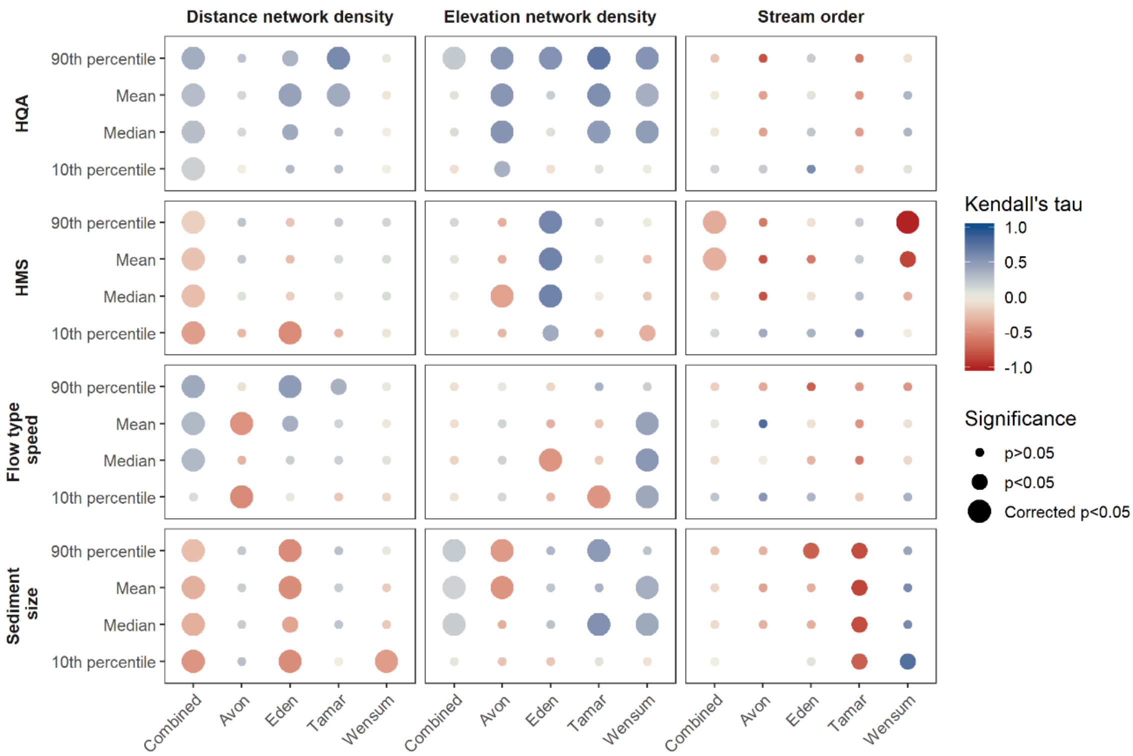


Figure 4.4. *Summary of correlations between distance network density, elevation network density, stream order and RHS variables for all catchments combined and each individual catchment. Significance of correlation is indicated by point size with the largest points significant post p-value correction. Effect size (Kendall's τ) is indicated by colour. No correlation was possible between stream order and 10th percentile sediment size in the Avon due to no variation in the RHS variable.*

There are numerous significant correlations between the network topology metrics and RHS variables for individual catchments, many of which were also shown to be significant after the correction to the p-value. The results show that catchments have different responses to the network topology metrics of distance network density and elevation network density. Distance network density only has significant correlations with the regulatory scoring variables (HQA and HMS) in the Eden and Tamar (**Figure 4.4**). Elevation network density, however, has a wider array of significant correlations with the scoring variables, particularly HQA, which shows subtle peaks and troughs reflecting the distribution of both network density metrics (**Figure 4.3a** and **b**). HMS shows mostly negative correlations, mainly with elevation network density, apart from the Eden, which has significant positive correlation across all HMS descriptive statistics (**Figure 4.4**). Visually, 10th percentile HMS is most variable to network density with peaks and troughs responding to the network density distributions (**Figure 4.3a** and **b**).

For individual RHS features, the response to network topology varies between catchments (**Figure 4.4**). The Avon has negative correlations between flow-type speed and distance network density, with an evident drop in 10th percentile flow-type speed associated with peaks in network density (**Figure 4.3a**). The Eden and Tamar, however, have positive correlations with mean and 90th percentile flow-type speed for distance network density but negative correlations with median and 10th percentile elevation network density. The Wensum shows positive correlations between flow-type speed and elevation network density. Sediment size has a consistent response to distance network density, with the Eden and Wensum showing negative correlations with the sediment size (**Figure 4.4**). For elevation network density, the Avon shows negative correlations with sediment size, whereas the Tamar and Wensum show positive correlations (**Figure 4.4**).

There were few significant correlations between stream order and the RHS variables in individual catchments (**Figure 4.4**). The only significant correlation after p-value correction is with 90th percentile HMS in the Wensum, which shows a strong negative relationship.

4.2.4 Discussion

4.2.4.1 A new approach to utilizing network topology in catchment-level analysis

Network density metrics represent an alternative approach to account for network topology in catchment-level studies, optimizing the width dimension of the network (or the number of links in the network) as opposed to the commonplace stream-order metric which only reflects the longitudinal position of links in a network (**Figure 4.1**). This study demonstrates that two topology metrics can be calculated simply from a DTM with GIS and, using a broad-scale analysis of river attributes, can be used to investigate the functional processes within catchments.

While the two network density metrics have similar forms (i.e. forms are consistently unimodal or multi-modal), the spatial configuration of the distance and elevation intervals used in the calculation of network density varies and may impact the effectiveness of each topological metric. Distance network density separates the catchment into intervals based on distance which spread upstream from the outlet (**Figure 4.2a**), reflecting natural sub-basins within the fractal structure of the catchment (Lashermes and Foufoula-Georgiou, 2007). This differs from elevation network density, which separates the catchment into intervals based on elevation which radiate out from the main channel of the network (**Figure 4.2b**). The configuration means that distance intervals contain streams that are in closer proximity to one another rather than the more distal configuration created by the elevation intervals, suggesting a degree of spatial dependency in river functioning. This has been highlighted in previous studies where spatial network structure has a stronger influence on some in-channel processes than predictor variables such as elevation (Steel *et al.*, 2016). However, elevation intervals contain RHS sites that, although they may be distal, may have similar properties as elevation has been strongly related to RHS variables including flow type, substrate, etc., in a number of studies (Jeffers, 1998a; Naura *et al.*, 2016; Vaughan *et al.*, 2013).

4.2.4.2 Impacts of network topology on river characteristics

River characteristics are assessed using the RHS dataset. The observations made by the RHS dataset (**Table 4.1**) cannot offer the level of detail regarding geomorphological process that river classifications which consider multiple scales can offer (e.g. Brierley and Fryirs, 2000; Gurnell *et al.*, 2016). While process-based classifications are preferable, broadscale monitoring datasets, such as the River Habitat Survey, may still be useful when combined with map-derived data to explore controls on river characteristics (Harvey *et al.*, 2008b; Naura *et al.*, 2016; Vaughan *et al.*, 2013). However, there are biases in RHS data collection, an inherent limitation when using existing datasets (Vaughan and Ormerod, 2010), specifically, the standardized survey length of 500m reach that will capture differing amounts of natural variability depending on the size of the river. While this must be noted, there are few significant correlations between river characteristics identified with stream order (**Figure 4.4**), which suggests that channel size is not influencing the RHS variables to a great degree in these catchments.

In this study, it is anticipated that sites in high network density areas will have higher levels of habitat diversity, as indicated by previous studies of confluences and networks (Benda *et al.*, 2004a; Best, 1985; Rice, 2017), in turn increasing mean sediment size and flow-type speed compared to sites in low-density areas. The results of the correlations between distance network density and river characteristics when all catchments are combined support this hypothesis, with greater HQA, flow-type speed and coarser sediment sizes observed in areas with high distance network density (**Figure 4.4**).

For individual catchments, elevation network density induces a stronger positive HQA response across all catchments than distance network density (**Figure 4.4**). This supports the evidence that individual confluences (Rice *et al.*, 2006) and high densities of confluences increase physical heterogeneity within the river network (Benda *et al.*, 2004b; Rice, 2017). However, flow-type speed and sediment size respond differently to network density in individual catchments.

Slower flow types are observed in high network density areas of the Avon and Tamar, whereas faster flow types are observed in high-elevation network density areas of the Wensum. Individual confluences are shown to create numerous high- and low-speed flow environments (Best, 1987) that may be observed in surface water topography (Biron *et al.*, 2002). It was expected that the introduction of the additional flow types by a high density of confluences in relatively low-energy rivers would increase mean reach flow-type speed; however, the correlation analysis (**Figure 4.4**) suggests that in some catchments mean flow-type speed is reduced.

Sediment size response also shows variation between catchments. Sediment size is coarser in network-dense areas of the Avon and Eden as expected, but is finer in both the Tamar and Wensum (**Figure 4.4**). The evidence from high-energy rivers shows that sediment becomes coarser downstream and finer upstream of certain confluences (Benda *et al.*, 2004a; Rice, 1998), and in this case high numbers of confluences were expected to increase mean sediment size of the reach. However, the impact on sediment size is dependent on the sediment calibre of the incoming tributary being higher than the main channel, with enough energy to transport the coarse sediment to the confluence for numerous tributaries in an area. The Tamar and Wensum have different ranges of sediment sizes, with the Tamar having coarser sediment than the Wensum on average (**Figure 4.3**). This implies that tributaries in the Tamar may be energy limited, not transporting coarse sediments to confluences, and the Wensum may suffer from inputs of fine sediments from the high percentage of arable land that concentrates in network-dense areas. This has before been observed in a low-energy modified catchment where anthropogenic modifications in tributaries reduced coarse sediment and flow capacity, causing either limited confluence impact or localized sediment fining (Singer, 2008).

Others have related the capacity of confluences to alter reach features to the morphometry of catchments, with larger and more circular catchments containing a higher percentage of confluences that have a significant impact (Rice, 2017). Based on this theory, the Eden and Tamar are likely to have the greatest impact as they are the most circular and steepest of the four catchments (although the Tamar is the smallest by area). Yet there is no clear pattern indicating that these catchments respond differently than the Avon and Wensum (**Figure 4.4**), with catchments responding differently to different variables. This perhaps suggests that rather than network density having a directional impact on factors such as flow type speed and sediment size, it has an impact on overall heterogeneity at the reach level, as suggested by previous studies, and that specific directional change occurs at the sub-reach level.

An increase in channel modification is also hypothesized due to the increase in flood peak downstream from confluences (Depettris *et al.*, 2000) and the scour and erosion associated with confluence junctions (Best, 1986) potentially increasing the need for bed and bank protection. The correlations between distance network density and HMS when all catchments are combined undermine this hypothesis, showing less channel modification where distance network density is higher (**Figure 4.4**). There were few significant correlations with individual catchments, but HMS in the Eden was consistently observed to be higher in network-dense areas (**Figure 4.4**). This may be due to the Eden's high elevation and steep topography inducing a higher energy environment where scour and erosion processes in areas with high numbers of confluences would be more likely to be present.

Differences in RHS variable responses also differed between the descriptive statistics considered. Often the extremes, 90th and 10th percentiles, showed more significant and stronger correlations than the mean or median (**Figure 4.4**). This may reflect findings from previous studies which suggest that not all confluences cause reach-scale changes (Rice, 1998), that perhaps the changes to river character induced by certain confluences only influence certain reaches, whereas others are left unaffected, creating less pronounced responses in the mean and median of variables. External factors may also influence this trend: for example, 10th percentile HMS visually responds dramatically to network density metrics (**Figure 4.3a** and **b**) compared to the other descriptive statistics.

This suggests that the most natural sites (i.e. with the lowest HMS score) are responding differently to network density, with the most natural sites having less modification in network-dense areas, whereas less natural sites become more modified in network-dense areas. This reflects the HQA score results which visually (**Figure 4.3a** and **b**) and statistically (**Figure 4.4**) vary with distance network density, except for the 10th percentile. These sites, with the lowest habitat quality and naturalness, are likely influenced by anthropogenic factors that are independent of network density, reducing habitat quality scores at impacted sites.

The two network density metrics have differing impacts on river characteristics. While distance network density shows consistently significant correlations when all catchments are combined, individual catchments respond more frequently and more strongly to elevation network density (**Figure 4.4**). This may be because there is a dramatic split in distance network density values between the more upland, drainage dense catchments, Eden and Tamar, than the lowland, chalk, low-drainage catchments, Avon and Wensum. The combined correlation will therefore in part reflect the difference between the catchments which have different ranges of river characteristics (**Figure 4.3**). This is not the case for elevation network density, which has higher density values in the Tamar and Wensum than the Avon and Eden, so therefore the characteristics of the catchments will have less bearing on the combined correlation. However, there are patterns identified with distance network density in individual catchments that are not present with elevation network density, increased flow-type speed and sediment size in the Eden and reduced flow-type speed in the Avon with network density (**Figure 4.4**), which show its usefulness.

The results suggest that the distance network density and elevation network density metrics quantify different dimensions of network topology which are shown to exhibit functionally meaningful patterns for river reach characteristics based on the correlation results. Perhaps within catchments elevation network density provides the more powerful metric for individual catchments, but distance network density better accounts for the drainage density of the catchments, allowing it to be applied across multiple catchments.

4.2.4.3 Comparison of stream order to network density metrics

The classic method of accounting for network topology, stream order, is critiqued for failing to represent discontinuities in the network and simplifying the network into a gradient. The number of links in different stream orders is consistent across all catchments not reflecting the internal structure of the network or the variety between catchments that is achieved by the distance network density and elevation network density metrics. The analysis of the two network density metrics presented in this paper shows that distances from source and elevation are not mutually exclusive (**Figure 4.2a** and **b**), contrary to the stream-order metric which represents streams as upstream to downstream or upland to lowland.

Stream order has few significant correlations with many of the river characteristics considered in this study. Negative correlations with HMS in the Wensum were statistically significant post p-value correction, likely driving the significant relationship with all catchments combined for this variable (**Figure 4.4**). This suggests that modification is greater upstream in the Wensum, contrary to ideas that downstream reaches may show greater anthropogenic modification. Intense agricultural land use in the upper reaches of the Wensum is likely to be the cause of the high HMS scores upstream. The lack of significant correlations suggests that stream order and therefore an upstream-to-downstream gradient are not the predominant pattern in river characteristics despite the description of such a gradient by geomorphic (Schumm, 1977) and ecological frameworks (Vannote *et al.*, 1980). This is surprising as distance and elevation, which both reflect the upstream-to-downstream gradient, have proven to be important factors in previous studies explaining patterns of RHS features at a national level (Jeffers, 1998a; Vaughan *et al.*, 2013). This implies that upstream-to-downstream gradient may not sufficiently reflect patterns of river characteristics through the river network within individual catchments. Others have also found that stream order has weak and inconsistent relationships with biodiversity patterns in river systems, arguing that the topological measure has no direct mechanistic control on biodiversity (Vander Vorste *et al.*, 2017). Instead, this study finds that the network density metrics are a more powerful metric which conceptually provide an improved method of accounting for the impacts of network topology on the fluvial system exhibiting relationships with river characteristics, particularly habitat quality score (**Figure 4.4**).

4.2.4.4 Applicability of network topology metrics to different environments

Much of the seminal work on network and confluence impacts (e.g. the Network Dynamics Hypothesis, Benda *et al.*, 2004b, and Link Discontinuity Concept, Rice *et al.*, 2001) is conducted in natural, highly erosive catchments with first-hand empirical measurements. However, in an age when rivers and their catchments are increasingly altered by anthropogenic modification (Meybeck, 2003), contemporary studies must not only aim to expand knowledge, but also find methods of transferring knowledge to many, increasingly altered, catchments (Clifford, 2002).

The catchments selected by this study are more greatly modified and, although they reflect a range of fluvial environments in England, are lower-energy catchments than the seminal studies. Benda *et al.* (2004a) suggest that confluence effects in less active landscapes would be subdued compared to highly erosive landscapes, but the evidence presented here demonstrates the utility of evaluating network topological structure in studies on catchment-level effects in any type of fluvial environment, including those with low-energy and widespread anthropogenic changes. The response of some river characteristics varied between catchments; observations of flow-type speed, sediment size and modifications showed different responses to network density in different catchments. This suggests that the functional effect of these topological metrics is catchment dependent and likely is influenced by external catchment characteristics such as land use not considered in this study, although impact did not appear to vary with catchment topography or circularity, as has been shown in prior studies (Benda *et al.*, 2004a; Rice, 2017). This should be explored further in future research to enable recommendations to be made regarding where and how reaches may respond to network density. The response of habitat quality score was, however, consistent across catchments and between metrics, showing that habitat quality is greater in areas with high network density (**Figure 4.4**), as hypothesized by the Network Dynamics Hypothesis (Benda *et al.*, 2004b) and demonstrated by studies on individual catchments (Rice, 2017; Rice *et al.*, 2006).

The methods presented in this paper are designed to be implemented in any catchment with a dendritic network structure. The topology metrics can easily be calculated from any dendritic network with DTM data using GIS and compared to any site-scale data. This study uses regulatory monitoring datasets so that analysis is targeted to assessment scores and physical features of interest to river managers. Also, the high volume and wide spatial extent of data available from regulatory sources allow for between-catchment comparisons.

4.2.5 Conclusions

Although appreciation of catchment-level effects is considered the epitome of understanding river functioning, a key component of the catchment – the river network – is overlooked and oversimplified by catchment-level studies. This study finds that river network density plays a role in structuring the distribution river characteristics throughout the catchment, offering more detailed explanation than the classic stream-order metric. Network-dense areas are generally found to have higher habitat quality and diversity, but modification, flow-type speed and sediment size show different responses in different catchments. This study suggests two potential reasons for this: (1) there is evidence that confluences in the river network increase diversity, as is observed in this study, so the direction of mean river characteristic response may not be consistent, and

(2) there may be external factors such as sediment availability, land cover and anthropogenic modification that alter the direction of mean river characteristic response. This paper demonstrates the functional response of river characteristics to network topology and suggests that the inclusion of network topology in catchment-level studies would add a layer of function-based understanding to such studies, linking reaches to their catchments.

The broad-scale methodology adopted by this study allows the network density metrics, which are easily extracted from open-source data using GIS software, to be compared to any regulatory dataset. The use of regulatory datasets allows not only for analysis over a wider spatial extent, but also for more applicable results for regulatory bodies. Therefore, the interdisciplinary approach to characterizing network topology can be applied efficiently and effectively to capture catchment-level impacts on reach-level functioning in any catchment across the globe.

Data availability: River Habitat Survey data are freely available from the Environment Agency, England. River network data and elevation data are available from the Centre for Ecology and Hydrology, UK.

Author contributions: EH and NC conceptualised and designed the study; EH performed the data analysis with the support of JM; EH wrote the paper with comments provided by NC and JM.

Competing interests: The authors declare that they have no conflict of interest.

Acknowledgements: The authors would like to thank the Environment Agency and the Centre for Ecology and Hydrology for access to the data used in this paper. This research is funded by the Natural Environmental Research Council (NERC).

4.3 SUPPLEMENTARY WORK

4.3.1 Network topology effects on physical habitat diversity indices

The physical habitat diversity indices (described in **Table 2.2**) are not included in Heasley *et al.* (2019) (Section 4.2) the as a response variable to focus the paper on the effects on habitat type and regulatory indices (**Table 4.1**). For consistency throughout the thesis the results of the diversity indices are presented in **Figure 4.5.**

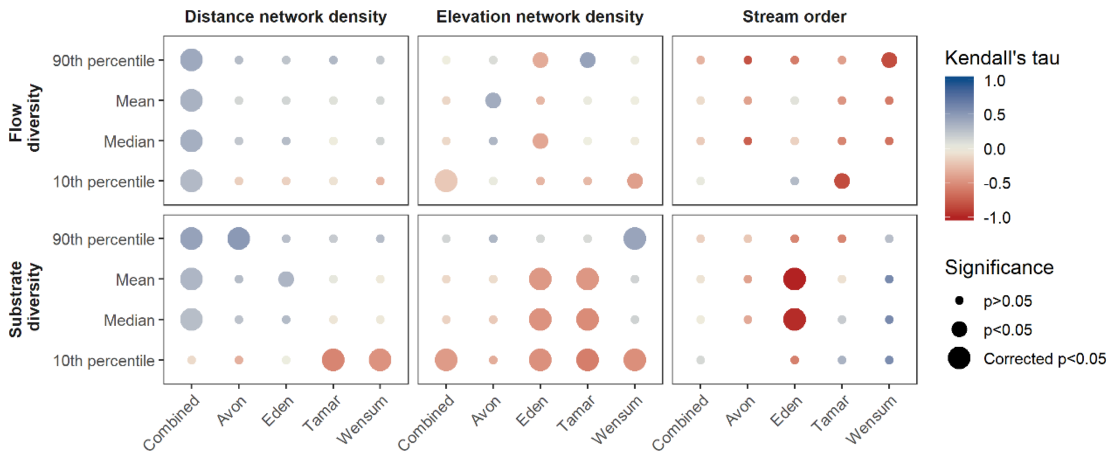


Figure 4.5. *Summary of correlations between distance network density, elevation network density, stream order and habitat diversity indices for all catchments combined and each individual catchment. Significance of correlation is indicated by point size with the largest points significant post p-value correction. Kendall's τ is indicated by colour. This figure is equivalent to* ***Figure 4.4*** *in the Section 4.2.*

The diversity indices show similar relationships with those habitat type and regulatory indices presented in the Section 4.2. The distance network density metric exhibits significant positive correlations with habitat diversity indices when all catchments are combined. However, there are fewer significant correlations with individual catchments. Elevation network density on the other hand shows more significant correlations with individual catchments, primarily negative correlations. This again highlights the differences between network metrics discussed in Section 4.2.4.2 and their applicability to explain differences in river characterises within and between catchments.

The differences between descriptive statistics are also consistent with the findings in Section 4.2. Extreme habitat values within each distance or elevation band (10th and 90th percentiles) have more significant correlations than the mean and median values. This may be because not all confluences in high density areas of the network influence reach features (Rice, 1998), influencing sites with the best and worst habitat conditions rather than mean habitat conditions in the area. **Figure 4.5** indicates that, for substrate diversity, the most heterogeneous sites (90th percentile) become more diverse at higher network densities whereas the most homogeneous sites (10th percentile) become less diverse indicating that high network density induces a greater range in habitat conditions in an area.

Also consistent with the Section 4.2 are the few correlations with stream order, with only negative correlations with mean and median substrate diversity present with corrected p-values <0.05.

The discussion in Section 4.2 stipulates that network density may not have directional impact on mean habitat type but has an impact on overall heterogeneity at the reach-level. This is because evidence from individual confluences shows that flow speeds may diversify due to this introduction of different flow environments at confluences (Best, 1987). However, there are few significant correlations between flow diversity and network density within individual catchments. Substrate diversity is correlated to elevation network density and generally becomes more homogeneous in network dense regions of some catchments, although may be more heterogeneous at sites with the greatest diversity. Elevation network density even exhibits significant negative relationships with mean and median substrate diversity for the Eden and Tamar catchments, effecting overall habitat conditions, rather than just certain sites. This may be because the Eden and Tamar are the most circular and steepest catchments and therefore likely have more confluences with a significant impact (Rice, 2017).

These results back up the findings from Heasley *et al.* (2019) in Section 4.2, highlighting again that distance network density may explain the effects of network topology regionally, whereas elevation network density may explain effects better in individual catchments. It also highlights how habitat diversity may not increase in network dense areas as expected but become more homogeneous. Only the sites with the greatest habitat conditions in the band exhibited increases in diversity with network density indicating how the effects of network topology are variable and complex.

4.4 CHAPTER CONCLUSIONS

Network topology is associated with river characteristics in four catchments in England. Metrics to quantify network topological structure (i.e. network density along distance and elevation gradients) that have their roots in flood hydrograph estimation are successfully adapted for application to any catchment with a dendritic network using a simple GIS procedure. The metrics successfully reflect the topological signature of each catchment, reflecting network shape and density even at a coarse resolution (i.e. 5% interval bands).

The network density metrics better describe river characteristics than stream order suggesting that a longitudinal up-to-downstream gradient in river condition to be an over simplified concept within eco-hydromorphological research. The topological structure of the network over distance best describes regional patterns of network density, with network dense areas having greater habitat quality and diversity in line with predictions from other studies (Benda *et al.*, 2004b; Rice, 2017). Elevation network density better describes longitudinal patterns of river characteristics within catchments, with different catchments showing different responses to the network density metrics which is likely due to differing catchment properties. River characteristics showed a functional response to network topology even in the low-energy landscape of England, compared to the high-energy landscapes where the seminal concepts on network structure were developed (Benda *et al.*, 2004b; Rice *et al.*, 2001). This highlights the utility in considering network topology within studies on catchment-level effects in any environment.

Network density is shown to influence certain RHS sites but is frequently not associated with mean or median river characteristics. However, network density was shown to influence extreme river characteristics within areas of the catchment. This is an interesting finding considering that broad-scale methodologies are designed to describe average trends in local values rather than the extremes. This suggests that high network density influences downstream patterns of river habitats in certain sites but neither network density nor stream order explain dominant downstream patterns in river habitats in these four catchments.

CHAPTER 5

Catchment and confluence properties influence the effect of network topology on river habitats

5.1 INTRODUCTION

Network topology has a long history of research (Wharton, 1994) and is an important and sensitive pathway of the catchment linking catchment-level effects to river habitats in downstream reaches. However, it is an overlooked aspect of the catchment system, being included in few studies of catchment controls on river systems (see evidence review in Section 2.1.3) and with little inclusion in applied river science literature. This is because many studies attribute more influence to catchment than local landscape characteristics (Loiselle *et al.*, 2016; Sliva and Williams, 2001), yet most do not explore how wider catchment influences are transported to river reaches (Allan, 2004). This work seeks to address this gap in literature by providing a better sense of where and how network topology affects river reaches to supplement understanding of catchment-level effects.

This chapter builds on results from Chapter 4 by exploring how features of the catchment and individual confluences may influence the effect of network topology on river habitats. Chapter 4 was based on the premise that longitudinal trends in reach characteristics reflect both the position of a reach along the network and the width of the network (summarised in this chapter in **Figure 5.1**).
This was expressed as two metrics of network density: (i) distance network density; and (ii) elevation network density. The metrics were calculated for four example catchments in England and network density was found to both reflect network topology and relate to habitat features. While there were relationships between network density and habitat feature extremes, the correlations developed with mean and median habitat features were not strong. The success of using network density to explain habitat features also varied between catchments. This chapter is motivated by the findings of Chapter 4 and investigates the processes that may explain why network topology has a positive, negative or no effect on physical habitats in different catchments.

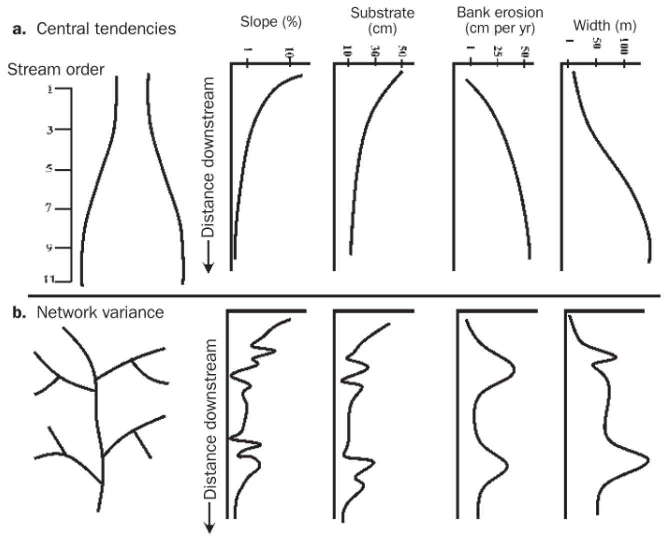


Figure 5.1. *Downstream changes in channel features from different perspectives: (a) the linear perspective of the river network with gradual downstream changes in reach characteristics (e.g. the river continuum concept; Vannote et al., 1980); (b) the network width perspective where the branching nature of the network disrupts downstream trends. Some features (e.g. slope and substrate) retain central tendency (e.g. the link discontinuity concept; Rice et al., 2001) whilst some remove it (e.g. bank erosion and width). Reprinted Figure 10 from Benda et al. (2004b, p. 424).*

The key assumption underlying the hypothesis that network density influences longitudinal patterns of habitats is that where there are more links in the network, and therefore more confluences, connecting multiple watersheds within a small area, reaches exhibit improved habitat condition (i.e. larger sediment sizes, more diverse flow types etc., see Section 5.1.1 for details). ***Important confluences*** refer to confluences that have a marked positive or negative effect on river condition from upstream of the confluence to downstream of the confluence. However, despite evidence that confluences are hotspots of diversity (see Section 5.1.1 for details), not all confluences have an important or a positive effect on physical habitats. Therefore, this chapter identifies which confluences are important and how important confluences may influence associations between network density and habitats. The number and distribution of important confluences in catchments is also informed by catchment morphometry which is explored in this chapter.

The three ways this chapter explores the effect of network topology on river habitats are summarised as three objectives:

Objective 5a: To identify how important confluences influence the effect of network topology on river habitats.

Objective 5b: To investigate which properties of upstream tributaries influence confluence importance.

Objective 5c: To explore how catchment morphometry influences the effect of network topology on river habitats.

Literature pertaining to each objective is explored in detail in Sections 5.1.1 to 5.1.3 and connections between objectives are summarised in **Figure 5.2**. The importance of network density is discussed in Chapter 4. Methods to address the objectives are presented in Section 5.2 and the results and discussion of each of the three objectives are provided in three individual sections (Sections 5.3, 5.4 and 5.5) with a discussion and conclusions presented in Section 5.6.

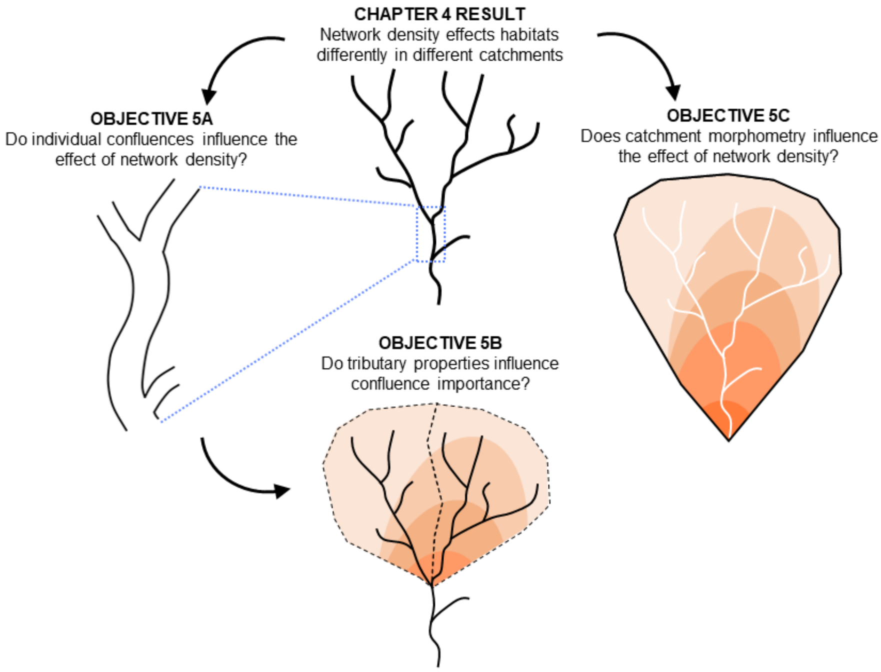


Figure 5.2. *Schematic of the key objectives of Chapter 5. Arrows indicate how Objectives 5a and 5c stem from the results of Chapter 4, and how Objective 5b is a natural progression from Objective 5a. Objectives 5a and 5b focus on individual confluences whereas Objective 5c focuses on properties of the catchment as a whole.*

5.1.1 The importance of confluences and tributary similarity

The hypothesis for Chapter 4 was that reaches in areas with a high density of links in the network (and therefore confluences) would exhibit greater habitat quality and diversity than reaches in low density areas of the network. This is because confluences in the network are repeatedly described as 'hotspots' of reach dynamism and diversity (McClain *et al.*, 2003) from both geomorphological and ecohydrological perspectives.

Geomorphological research shows that confluences create discontinuities in attributes such as sediment size, channel dimensions, slope and bars at confluences (Benda *et al.*, 2004a; Ferguson *et al.*, 2006; Rice and Church, 1998; Swanson and Meyer, 2014). This also creates a range of hydraulic environments and surface topographies (Best, 1987; Biron *et al.*, 2002) (**Figure 5.3**). The altered flow velocity, substrate and morphology around the confluence, produce greater dynamism and diversity, improving physical habitat availability. Thus ecological research also shows changes at confluences, for example, there are reported increases in factors such as riparian complexity, large woody debris and nutrient concentrations, along with increases in fish and invertebrate species richness and diversity at confluences (Fernandes, 2004; Kiffney *et al.*, 2006; Rice *et al.*, 2006; White *et al.*, 2018).

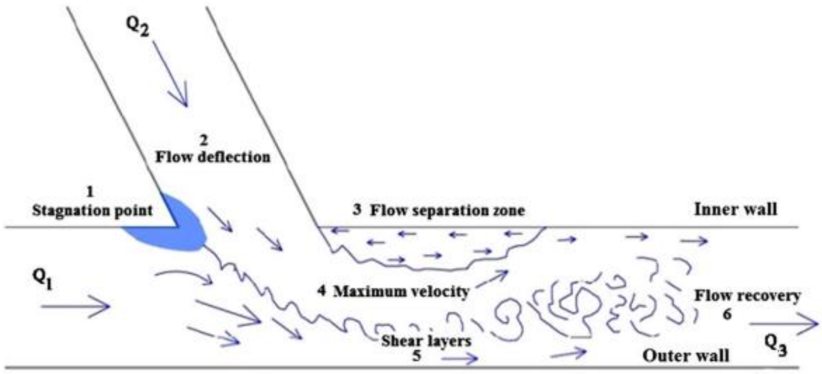


Figure 5.3. *Flow dynamics at a channel confluence creates diverse flow, morphology and substrate conditions. Reprinted Figure 1 by Leite Ribeiro et al. (2012, p.2), originally modified from Best (1987).*

Confluences may also be sites of modification as flooding may occur downstream of large confluences as flood peaks converge (Depettris *et al.*, 2000) potentially leading to more modifications for flood protection around confluences. Similarly, there is also increased scour and erosion at confluence junctions (Best, 1986) potentially leading to more bank protection around confluences.

However, most confluences have no important effect on river reaches, with only 14% of tributaries found to have an important effect on bedload size (Rice, 1998), which may alter the effect of network density metrics on river habitats. Important confluences (often known in the literature as 'significant confluences') have a marked impact on form and/or function at or just downstream of the confluence compared to conditions in surrounding channels. Geomorphic studies have focused on the difference in hydromorphological properties of the two incoming tributaries to explain confluence importance. Hydromorphological properties of the tributaries include discharge, bedload flux and bedload grainsize (Best, 1987; Ferguson *et al.*, 2006; Rhoads, 1987; Rice *et al.*, 2006) and surrogate measures such as Shreve magnitude, catchment area, slope and the product of area and slope (Benda *et al.*, 2004a; Rice, 1998; Richards, 1980).

The ratio between the hydromorphology of the secondary tributary (the smaller, incoming tributary) to the primary tributary (the larger, mainstem tributary) is often found to control confluence importance, with equally sized tributaries (in terms of discharge, upstream area or magnitude) found to have the greatest effect on riverine geomorphology, flow dynamics (Benda *et al.*, 2004a; Best, 1987) and ecological communities (Jones and Schmidt, 2018). The size ratio is independent of scale, suggesting that a confluence of two equally small headwater streams will have an important effect, as will two equally large downstream channels. However, there are studies that indicate that size similarity does not always yield positive effects. For example, Osborne and Wiley (1992) found more fish species in small tributaries that connected to the large main channel, than in similarly sized small headwater streams.

Geomorphological research has also highlighted that the stream power of individual tributaries is important, as both the total stream power for the secondary tributary and the ratio of catchment area between tributaries are successfully used to predict confluence importance (Rice, 1998). Similarly, the effect of size ratios becomes less important as slope of the primary tributary increases (Khosravinia *et al.*, 2019).

As important confluences likely alter the influence of network density metrics on river habitats (as network density measures the density of all confluences not just those with an important effect), it is appropriate to determine why some confluences have an important effect on habitats, while others do not. Most research into confluence importance investigate impacts on reach geomorphology and flow hydraulics, and therefore focuses on hydromorphological properties that reflect sediment and flow regimes. Most of these studies conclude that tributary similarity produces important confluences, however, there are other tributary properties that influence a wider variety of reach characteristics as now discussed in Section 5.1.2.

5.1.2 Wider catchment properties and tributary dissimilarity

While most previous research focuses on hydromorphological features of tributaries as they enter the confluence, hydromorphological features do reflect processes in the wider catchment. This is noted in network studies, for example, sediment load and grainsize availability of individual tributaries is determined by underlying geology (Knighton, 1980) or land cover influences (Jacobson *et al.*, 2001). The confinement of the valley also controls the effect a confluence can have when there is less opportunity for lateral adjustment (Swanson and Meyer, 2014).

The convergence of dissimilar watersheds at confluences may also increase confluence importance, but this has not been considered by the main geomorphological theories. Examples of dissimilar confluences may be seen at large rivers converging around the globe that cross major geological and ecological boundaries (Jones and Schmidt, 2016; **Figure 5.4a**) but also at smaller scales, where tributaries have differing properties such as land cover (**Figure 5.4b**).

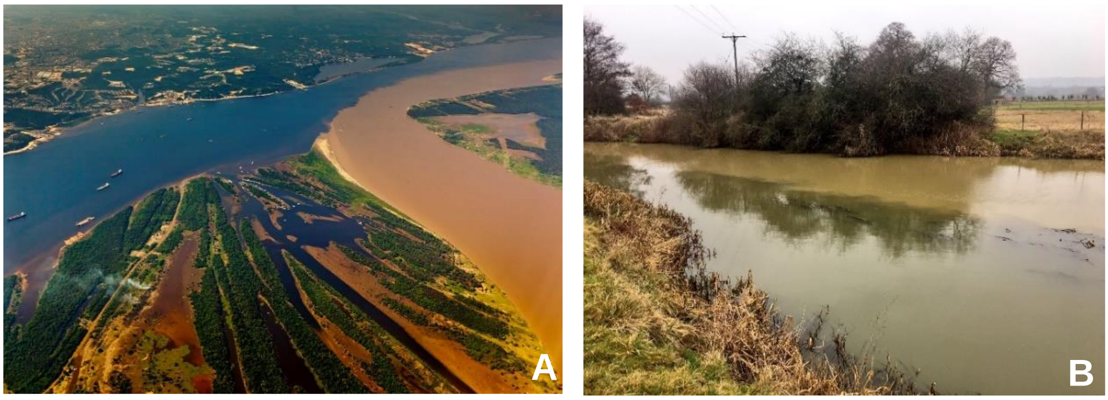


Figure 5.4. *Examples of tributaries with different catchment characteristics: (a) Encontro das Águas – Meeting of Waters – confluence in the Amazon basin where tributaries have differing geologies influencing sediment load* (Park and Latrubesse, 2015). *(Source: By Portal da Copa, CC BY 3.0, https://commons.wikimedia.org/w/index.php?curid=53416910); (b) River Beult and River Teise confluence, Kent UK, where one agriculturally dominated tributary increases the fine sediment load downstream (Source: Jay Neale, per comms).*

Dissimilarity between tributaries is not only reflected by the sediment load, but also in the ecology. Fish species abundances and diversity are found to be greatest when a tributary that experiences high flow disturbance converges with a stable main channel (Boddy *et al.*, 2019) and the convergence of contrasting stream types have the potential to create ecotones (Jones and Schmidt, 2016).

Dissimilarity of catchment characteristics (including baseflow index, and percentage of wetland, forest, agriculture and clay geology in the upstream watershed of each tributary) is included in Jones and Schmidt's (2016) model to predict which confluences will have the greatest impact on stream ecology. Confluences with equally sized tributaries and highly different catchment characteristics have the highest probability of having an important effect. The model has been applied to a watershed in Ontario where 13% of confluences are estimated to be important for ecology, however, this prediction was not tested against empirical data.

Most empirical research on confluences has been conducted in high-energy streams in north-west US and Canada (see review by Benda *et al.*; 2004a). There are few empirical studies on low-energy or less natural systems. Those that have studied such environments report a negligible influence of tributaries on downstream sediment sorting trends (Singer, 2008). Negative impacts at confluences have been observed when a tributary with poor water quality joins a high-quality stream, reducing the presence of freshwater mussels (Cooksley *et al.*, 2012). Therefore, important confluences may not always be hotspots of biodiversity (McClain *et al.*, 2003) but can alternatively be hotspots of negative catchment effects. It is therefore necessary to identify under which conditions confluences create positive or negative effects on habitats downstream.

In the UK, there are only a few studies focusing on tributary impact. They report step-changes in mean annual flood peaks at large confluences (Knighton, 1987) and the high density of several large confluences has been shown to push maximum stream power further downstream following increases in slope and discharge (Knighton, 1999) in the River Trent. There is therefore a need to explore the impacts of network topology in catchments with different characteristics to identify where network effects are most prominent.

5.1.3 Catchment morphometry and the Network Dynamic Hypothesis

The wider topological structure of the network is controlled in part by catchment morphometry which may control the number and location of important confluences within a catchment. By identifying a relationship between how network density metrics affect river habitats in individual catchments and the catchment's morphometry, it may be possible to explain why different catchments show different habitat responses to network density.

The Network Dynamic Hypothesis (NDH) proposed by Benda *et al.* (2004b), suggested that catchment morphometry influences confluence importance, and therefore likely impacts the effect of network density on river habitats. The NDH comprises several predictions about the likelihood of geomorphically important confluences in relation to network structure, catchment morphometry and watershed disturbances. Two of these hypotheses lead in part to the hypothesis that areas of the catchment with higher network density (and therefore more confluences) would have greater habitat quality and condition: (i) valley segments with closely spaced tributaries will have higher physical heterogeneity, compared to those with sparsely spaced confluences; and (ii) catchments with higher drainage density (and therefore higher topographic roughness) will have a higher degree of morphological heterogeneity.

Another NDH hypothesis may explain why network density has an impact in some catchments rather than others: compact catchments containing dendritic networks have more equally sized tributaries and hence more important confluences, compared with elongated basins containing trellis or parallel networks (**Figure 5.5**). Equally sized tributaries with smaller upstream areas are more common that equally sized large tributaries, suggesting that there may be more important confluences upstream (**Figure 5.5**). The NDH assumes that equally sized tributaries produce important confluences, however, Rice (2017) also finds that important confluences (defined as tributaries with similar size and similar power) respond consistently with the NDH to catchment shape.

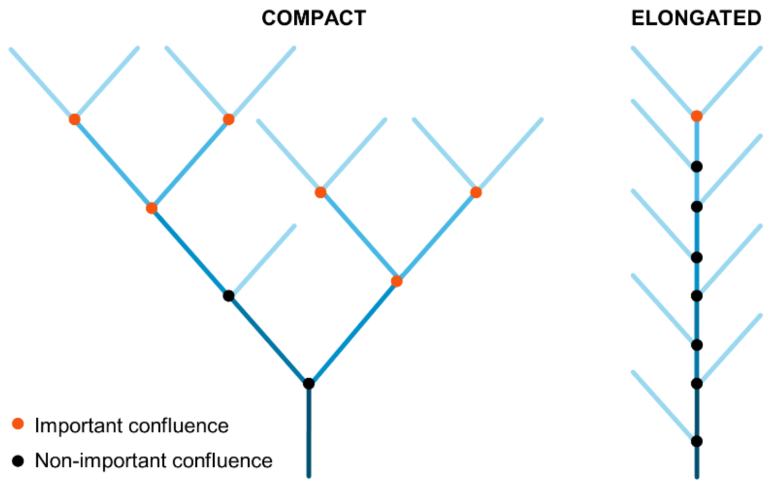


Figure 5.5. *Compact networks in circular catchments contain more important confluences, with equally sized tributaries, than elongated networks where important tributaries are limited to upstream. Prediction from the NDH* (Benda *et al.*, 2004b).

The topography of the catchment may also influence the effect of network topology on reach features, as coarse sediment supply and transport capacity are likely to influence confluence importance (Ferguson *et al.*, 2006; Rice, 2017). Upland catchments in the UK are more likely to have a coarser sediment supply because of harder geology and higher coupling to hillslope processes. Steep slopes and high rainfall also increase sediment transport capacity in upland areas (Raven *et al.*, 2010). Therefore, it is expected that habitats in upland catchments may respond more strongly to network topology.

5.2 METHODS

5.2.1 Network density metric calculation across England

To determine how the influence of network topology on river habitats varies with confluence and catchment properties, the network density metrics developed in Chapter 4 are used again in this analysis (see Chapter 4.2.2 for details of methods). The methods used to derive the network metrics are described again in brief below for convenience.

The same dataset is used as in Chapter 4, the Centre for Ecology and Hydrology's 1:50,000 blue-line network (CEH, no date-a; Moore *et al.*, 1994) to calculate the metrics. A dendritic network with no anabranches is essential for network metrics calculation (Section 4.2.2.2) so a protocol for removing anabranches for the entire network of England was implemented (see **Appendix 5A** for description).

Individual networks were identified within each catchment boundary (Chapter 3.3.2.1 for details of catchment boundary creation) using the RivEX add-on to ArcGIS v10.3 (Hornby, 2010). Only catchments with a dendritic network were selected for analysis, leaving 703 catchments.

Linear metrics of distance from the outlet and elevation above the outlet were extracted from the network data using RivEX and from the DTM elevation dataset used in Chapter 4 (CEH, no date-b) respectively. The linear metrics reflect the downstream gradient in reach features commonly used to represent network topology (**Figure 5.1a**) to compare to network density metrics that reflect network width (**Figure 5.1b**).

Network density metrics were calculated by splitting each network into distance and elevation bands with boundaries drawn at intervals of 5% of distance from outlet or elevation above outlet. The number of links within each band was calculated and divided by the width of the band to reflect network density over distance or elevation. Linear metrics and network density metrics were normalised between zero and one within each catchment so analysis focuses on within catchment variation in network topology rather than picking up on regional patterns (see Chapter 4.2.5). The code for calculating the network density metrics across England is presented in **Appendix 5B**.

Error in the network density metrics is possible as the networks are partially created from a DTM which has error associated with it. Network width function, the basis of the distance network density metric, is highly sensitive to elevation error, especially further from the outlet, as low magnitude links are often difficult to map (Lindsay and Evans, 2008). The network and its confluences are considered static for the purposes of this research. However, the expansion of the perennial network (Wharton, 1994) will create temporal fluctuations in network density furthest from the outlet, although this is unlikely to alter the broad topological signal of catchments which is the focus of this work.

5.2.2 Confluence importance

River Habitat Survey (RHS) data were used to identify confluences with an important effect on physical habitats. Typically the effect of a tributary is assessed by comparing features found on the main channel upstream of a confluence to those downstream of a confluence (Jones and Schmidt, 2018; Rice *et al.*, 2006 etc.). This is the approach employed here. There are 853 confluences in England where there is an RHS site both on the primary tributary and within 500m of the confluence downstream (**Figure 5.6**).

Confluence importance was assessed in two ways:

- ***Confluence effect*** is the percentage change in the habitat index downstream of the confluence, positive for an increase or negative for a decrease in habitat index value.
- ***Confluence strength*** is the percentage change in the habitat index downstream of the confluence irrespective of directional change, so larger numbers indicate stronger effect, and lower numbers weaker effect.

Confluence strength and effect were calculated for the habitat indices used throughout this thesis: two summary indices (HQA and HMS) and four habitat indices (sediment size, sediment diversity, flow type speed and flow type diversity) (see **Table 2.2** for description). Higher values of HQA, diversity, flow type speed and sediment size, and low values of HMS indicate better habitat condition.

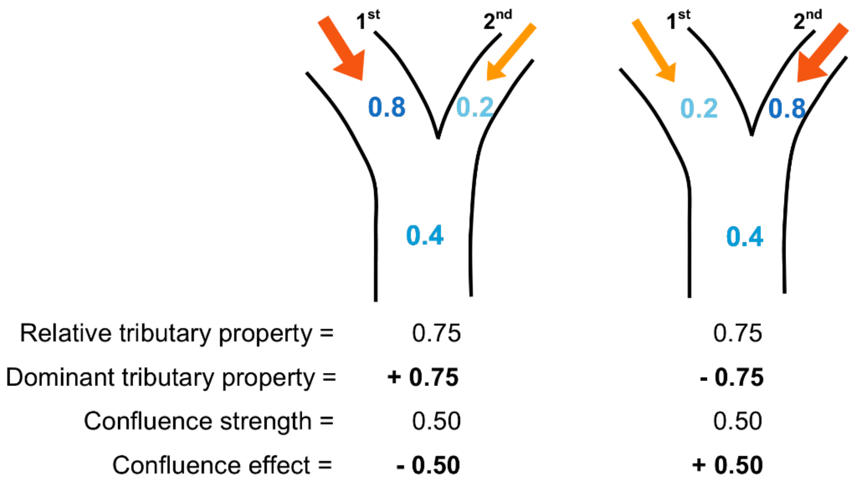


Figure 5.6. *Schematic of confluence importance and tributary properties measures. Confluence importance for habitat indices (indicated by blue numbers) can be similar for confluence strength but differ for confluence effect. Similarly, tributary properties (indicated by arrows) may have similar relative properties but different dominant properties (positive if the primary tributary is dominant, and negative if the secondary tributary is dominant).*

5.2.3 Tributary properties

Properties of primary and secondary tributaries (**Table 5.1**) are compared to explore which upstream factors influence confluence importance. Stream magnitude is used as a proxy for discharge (Knighton, 1998), because catchment area is too time consuming to calculate accurately and quality control at a national level (Pryde *et al.*, 2007), whereas stream magnitude is not dependent on DEM resolution and can be extracted easily in RivEX. The tributary with the largest magnitude is classified as the primary tributary.

Slope reflects the energy of each incoming tributary and is calculated from the DEM in ArcGIS, a method used by other studies exploring catchment effects (Leal *et al.*, 2016; Manfrin *et al.*, 2016). It is extracted both locally (at downstream end of each tributary as it enters the confluence) and as mean upstream slope within a 100m buffer of each upstream tributary. Percentage of each geological and land cover category are extracted within a 100m buffer of each tributary (see Section 3.2.2.1 for description of catchment datasets and **Appendix 5C** for detailed description of method).

Dissimilarity between the primary and secondary tributaries is calculated by two methods:

- ***Relative tributary properties*** are calculated as either absolute differences, for the categorical stream magnitude, geology and land cover properties, or as a ratio between the primary and secondary tributaries for the slope variables (**Table 5.1**). Low values indicate similar tributaries while large values indicate dissimilar tributaries.
- ***Dominant tributary properties*** are similar to relative tributary properties, except values are positive when the primary tributary has a larger value of the tributary property, or negative when the secondary tributary has a larger value.

The differences are described in a hypothetical example in **Figure 5.6**. Confluence strength is compared to relative tributary properties whereas confluence effect is compared to dominant tributary properties using statistics (Section 5.2.5). Therefore, how tributary properties affect the magnitude and direction of confluence importance can be identified.

Table 5.1. *Tributary properties used in this study.*

Hydromorphology properties	Geology properties	Land cover properties
• Stream magnitude • Local slope • Mean upstream slope	• Hard rock geology • Other limestone • Chalk • Sandstone • Other sedimentary rock	• Urban • Arable • Improved grassland • Semi-natural grassland • Mountain/heath/bog • Woodland

5.2.4 Catchment morphometry

Two measures of catchment morphometry that are reported to influence the importance of confluences (see Section 5.1.3 for details) are used to explore why network density only influences habitat in certain catchments: circularity ratio, to reflect the shape of the catchment and; mean elevation, to reflect the energy of the catchment.

Circularity ratio represents the shape of the catchment which influences the distribution of important confluences (**Figure 5.5**) and can easily be extracted from a GIS. Circularity ratio (C) is defined by Miller (1953) as the ratio of the basin area of a circle having a circumference equal to the perimeter of the basin:

$$C = \frac{4\pi \cdot A}{P^2}$$

(Equation 5.1)

where A is waterbody area (km²) and P is the perimeter of the waterbody (km). As circularity ratio approaches one the catchment is more circular whereas when it approaches zero the catchment is more elongated (Stepinski and Stepinski, 2005).

Mean elevation of each catchment is used to represent the broad differences in morphological processes between upland and lowland catchments. This is because confluence importance is influenced by coarse sediment supply and transport capacity which tends to be greater in upland catchments in the UK (Raven *et al.*, 2010). Mean elevation is calculated within each catchment boundary using ArcGIS.

5.2.5 Statistical analyses

5.2.5.1 Isolating network topology influence from linear gradients

Network density metrics are significantly ($p<0.01$) correlated with the linear metrics although correlations are relatively weak (distance, $\tau=0.24$; elevation, $\tau=-0.08$). A first run of the analysis showed that many catchments with significant correlations between network metrics and habitat features, also have significant correlations with linear metrics. This makes it impossible in many catchments to isolate the effect of network density (**Figure 5.1b**) from the linear gradient (**Figure 5.1a**). Therefore, catchments with a significant correlation ($p<0.05$ after correction to the p-value, discussed below) between network density and linear metrics are removed from further analysis (628 catchments were retained). After these catchments are removed, 602 and 560 confluences remain in catchments that have no correlations for the distance and elevation metrics respectively.

5.2.5.2 Kendall's τ correlations and regressions

Correlations are conducted between linear metrics, network density metrics, and habitat indices following the methodology employed in Chapter 4, to identify which catchments have a significant correlation between habitat indices and network density. Kendall's τ correlation method is used as the variables are non-parametric and contain ties. The direction and strength of the correlation is determined by τ. False Discovery Rate (FDR) correction is applied to the p-values when a high number of correlations is conducted to control for false positive results.

Additional Kendall's τ correlations are used to determine the association between confluence importance and upstream tributary properties, and between catchment morphometry and the strength of the correlation (τ) between network density and habitat features.

Linear regression is used to determine the relationship between confluence importance and network density so that regression lines may be plotted.

5.3 NETWORK TOPOLOGY IMPACT ON HABITAT INDICES

This section presents the results and discussion for Objective 5a: To identify how important confluences influence the effect of network topology on river habitats. First, the effect of network topology on river habitats is explored for catchments in England (Section 5.3.1). Then, the effects of individual confluences on river habitats are assessed (Section 5.3.2), before these strands are combined to address the objective (Section 5.3.3).

5.3.1 Correlations between network topology metrics and habitat indices in England

The number of significant correlations between habitat indices, the network density metrics (**Figure 5.1b**) and the linear metrics (**Figure 5.1a**) were relatively low (**Figure 5.7**). FDR correction is applied to control for the probability of false positives in the results, however, this likely also introduces false negatives so significant correlations with and without FDR correction are considered.

Many catchments are removed (n=75) as they had correlations between network density and linear metrics (**Figure 5.7**) to isolate the effect of network density from the background downstream gradient. This indicates that both features of the network (density and downstream gradient) may influence habitat features as, for example, Milesi and Melo (2013) found greater effects of confluences on macroinvertebrate assemblages in peripheral regions of the catchment.

5.3.1.1 Correlations between linear network metrics and habitat indices

Prior to the removal of catchments where linear and density metrics were significantly correlated, 25-31% and 23-32% of catchments exhibited significant correlations between habitat indices and distance and elevation linear metrics respectively. However, after catchments were removed this number declined to 13-20% and 12-21% of catchments that exhibited significant correlations respectively. The low number of significant correlations between the linear metrics and habitat indices is surprising as distance and elevation have been shown to be important factors in previous studies explaining patterns of RHS features at a national level (Jeffers, 1998a; Vaughan *et al.*, 2013). This may be because this study explores within-catchment habitat relationships, whereas previous studies explore relationships at a national level which capture a greater range of distance and elevation values that are partially reflective of other national regional controls such as geology, climate and land cover.

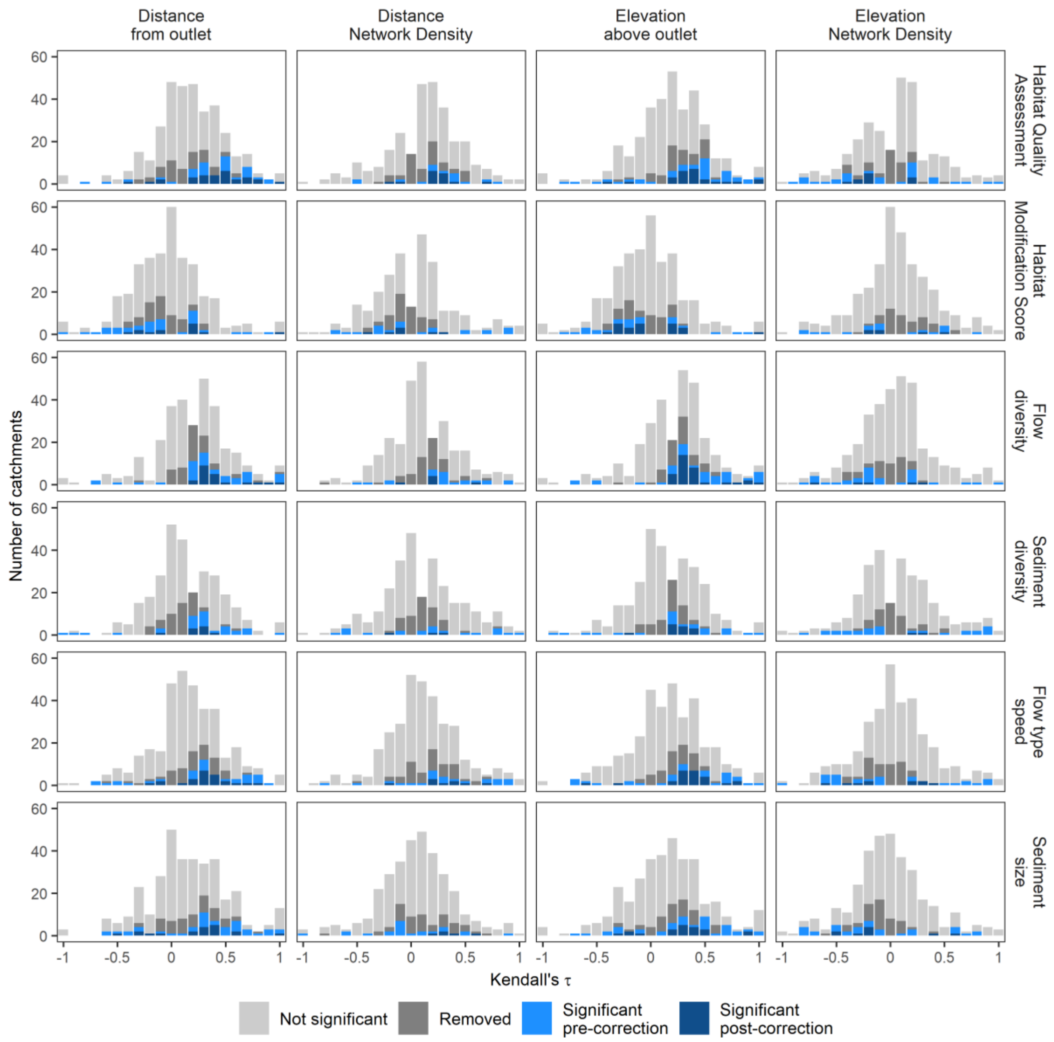


Figure 5.7. *Stacked bar chart indicating the number of catchments of with correlations between network metrics and habitat indices. Strength of correlation indicated by Kendall's tau and significant correlations are indicated (p<0.05) pre- and post-FDR correction to the p-value. The removed classification indicates the number of catchments with significant (p<0.05) correlations between the network density and linear metric.*

Significant correlations between the linear metrics and habitat indices were frequently positive, except for HMS, indicating better habitat conditions upstream (**Figure 5.7**). This reflects expected longitudinal patterns from continuum theories of river features (e.g. River Continuum Concept; Vannote *et al.*, 1980) and river processes (e.g. Process Domains Concept; Montgomery, 1999; and hillslope coupling; Church, 2002). The continuum theories propose that gradients in slope, shading, temperatures, nutrients, processes and channel dimensions create upstream environments with a broad range of habitats, coarser sediment, faster flows and little modification which changes gradually downstream. It must be noted that diversity indices are likely to be higher in smaller, upstream channels because of RHS sampling strategy where a 500m RHS reach captures a wider range of diversity in a small river than a larger river.

There were some catchments that do not follow the downstream gradient, particularly for the sediment size index, that have lower habitat conditions (e.g. finer sediments) upstream than downstream (**Figure 5.7**). This may be because the elevation profile in UK catchments may not always follow the profile in which many of the linear theories were developed, i.e. steep bedrock headwater streams with narrow valleys upstream with high coupling to hillslope processes. Instead, many catchments in the UK have an upstream plateau typical of upland moors, heathlands and grasslands which have deep soil horizons and lower slopes. These upland landscapes are comparatively uncoupled from steep hillslope processes which may reduce the capacity of reaches to exhibit the coarsest sediment sizes and fastest flow speeds. Sediment size may also be finer upstream because of factors such as a change in sediment supply (e.g. increases in fine sediment from agricultural practices) and human modification (e.g. straightening rivers, bank and bed protection and online structures such as weirs, sluices etc.) that alter sediment regimes and increase habitat homogenisation (Sear *et al.*, 2003; Walsh *et al.*, 2005).

5.3.1.2 Correlations between network density metrics and habitat indices

After the removal of catchments where linear and density metrics were significantly correlated, there were fewer significant correlations between habitat indices and the network density metrics than linear metrics. Only 8-11% and 7-14% of catchments analysed exhibited significant correlations between habitat indices and distance network density and elevation network density respectively (**Figure 5.7**). It is not surprising, given the relatively few correlations with the linear metrics, that the network density metrics had few significant correlations with habitat indices. It suggests that other factors are more likely to have stronger associations with habitat indices than the reaches' position in the network or the local network topology. For example, Singer's (2008) study of a lowland, low sediment supply river (Sacramento River, CA), showed that tributaries had a negligible effect on downstream fining rates. It is therefore expected that a high density of tributaries may have less association with river features in the comparatively gentle topographic landscape of England, than the steep, high-energy environments of north-west America where much of the seminal work on tributary effects has been conducted (Benda *et al.*, 2004b; Rice *et al.*, 2001). It will therefore interesting to explore the properties of confluences and catchments where network density is related to RHS features (see Sections 5.33 and 5.5).

Catchments that had significant correlations with one network density metric (either distance network density or elevation network density) frequently had a significant correlation with the other network density metric. However, the direction of association with habitat indices often differed between the two metrics. This suggests that catchments with significant correlations have a property that promotes a stronger association between network density and river habitats, rather different network density metrics being influential in different catchments.

Distance network density primarily displayed positive correlations with habitat indices, except for HMS which is negative (**Figure 5.7**). This supports the expectation that higher network density would have better habitat condition as individual confluences create increased flow speeds, sediment sizes and habitat diversity (see Section 5.1.1). Conversely, elevation network density primarily displayed negative correlations with habitat indices, except for HMS which is positive (**Figure 5.7**). However, within all habitat indices there are exceptions to these broad patterns (**Figure 5.7**) likely because of modifications and natural variation not accounted for in this analysis (discussed previously in Section 5.3.1.1).

5.3.2 The effect of confluences on river habitats

The effect of confluences on river habitats was broadly assessed by comparing RHS sites near confluences (<500m downstream from a confluences) to RHS sites far from confluences (>500m downstream from a confluences). There were significant differences between sites near confluences and far from confluences for five of the six habitat indices considered (Mann Whitney U, $p<0.01$). Only sediment diversity showed no significant difference between sites near and far from confluences (Mann Whitney U, $p=0.53$) despite others finding greater sediment size diversity downstream of confluences (Kiffney *et al.*, 2006). HQA and flow diversity was found to be higher, flow type speed slightly faster, and sediment size coarser nearer to confluences whereas HMS was lower nearer to confluences (**Figure 5.8**).

Individual confluence importance (i.e. the positive or negative effect on river condition downstream of the confluence) was assessed by comparing habitat indices upstream of the confluence to indices downstream of the confluences. Confluence importance was relatively symmetrically distributed for most indices with a relatively equal amount of positive and negative habitat effects downstream of the confluence (**Figure 5.9**). The diversity habitat indices showed that frequently an upstream site with no diversity had a downstream site with the highest diversity or vice versa. HMS on the other hand strongly increased downstream of confluences. The distribution of flow type speed was slightly negatively skewed indicating that flow type speeds tend to be slightly greater downstream of a confluence compared to upstream.

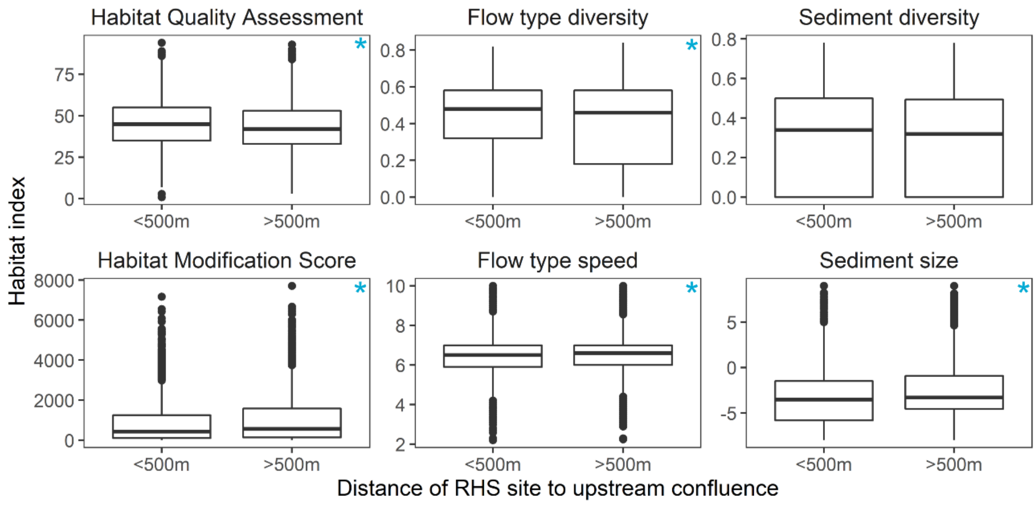


Figure 5.8. *Distribution of RHS variable scores for all RHS sites near confluences (<500m) and far from confluences (>500m).*

HQA, flow type speed and sediment size show that most confluences studied (23-33% of all confluences) had an important effect on downstream habitat (i.e. there was over 25% change in habitat index downstream of the confluence) which is slightly greater than previous studies that report ~13% important confluences (Jones and Schmidt, 2016; Rice, 1998). HMS and the diversity indices on the other hand showed a much higher percentage (55-80%) of important confluences (**Figure 5.9**). The calculation of confluence importance from broad-scale monitoring datasets has not previously been attempted. Although the sampling strategy of the RHS was not designed to explore confluence effects, it is possible to use these data to explore changes between habitats upstream and habitats within 500m downstream of 853 confluences. Consequently, this is one of the largest studies on confluences conducted ever using empirical data (Benda *et al.* 2004a used 167 confluences from 14 studies), covering a range of different environments. This highlights the benefits of utilising monitoring datasets for broad-scale scientific research.

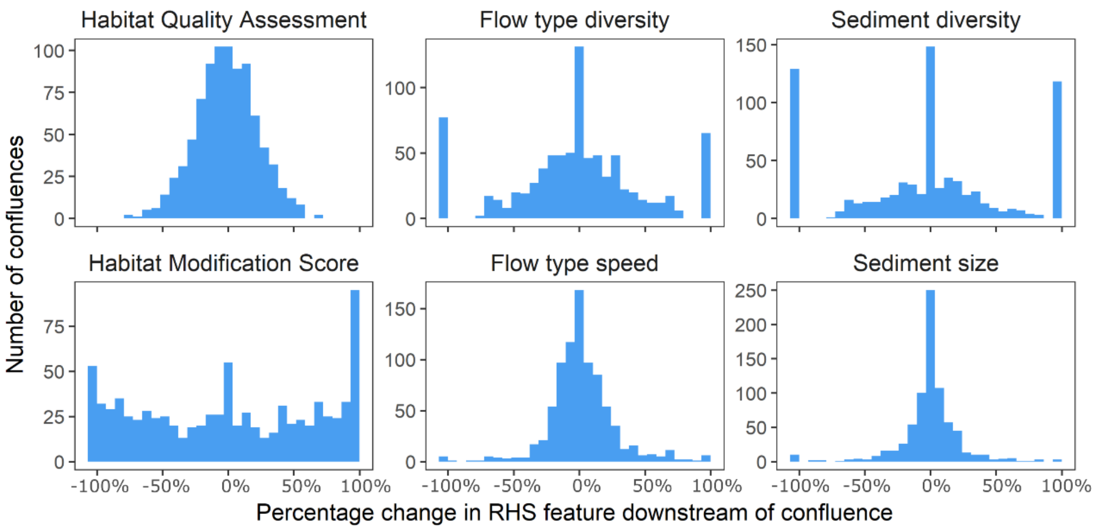


Figure 5.9. *Distribution of percentage change in habitat features downstream of the confluence. A value of 0 indicates no change downstream.*

5.3.3 Important confluences influence relationships between network density and habitat indices

Based on the correlations between network density metrics and habitat indices in **Figure 5.7**, three types of catchment response are identified: (i) ***Positive catchments***, which exhibited a significant positive correlation between a network density metric and the habitat index; (ii) ***Negative catchments***, which exhibited a significant negative correlation between a network density metric and the habitat index; and (iii) ***Non-significant catchments***, which exhibited no significant correlation between a network density metric and the habitat index in question.

The presence of important confluences in network dense areas is explored in relation to the type of catchment response. Separate linear regressions between network density and confluence effect are conducted for positive and negative catchments (**Figure 5.10**). Non-significant catchments are not assessed.

The sparse distribution of points in **Figure 5.10** indicates that confluence effects on habitat indices vary widely within both high- and low-density regions of their respective catchments. However, when split into positive and negative catchments the regression indicates some associations which exhibit significant linear relationships.

It is anticipated that positive catchments would have more positive confluence effects at higher densities, and negative confluence effects in low network density areas. This is because of the hypothesis from Chapter 4, that high network density would improve habitat condition at a broad scale due to cumulative improvements from individual important confluences (see Section 5.1.1 for summary of positive habitat improvements at confluences). However, many negative catchments are also observed (**Figure 5.7**). In negative catchments, it is anticipated that there would be more negative confluence effects at higher densities, and positive confluence effects in low network density areas, that would drive the negative associations between network density and the habitat indices.

Sediment diversity and flow type speed support this hypothesis as confluence effects become more positive with distance network density (sediment diversity, $p=0.01$, $R^2=0.06$; flow type speed, $p=0.03$, $R^2=0.02$) and elevation network density (sediment diversity, $p=0.01$, $R^2=0.10$; flow type speed, $p=0.07$, $R^2=0.03$) in positive catchments (**Figure 5.10a**). For elevation network density, sediment size also supports the hypothesis visually, with negative confluence effects in high density areas of negative catchments. Confluence effect on HMS also becomes more visually positive in high density areas in positive catchments and more negative in high density areas in negative catchments (**Figure 5.10b**).

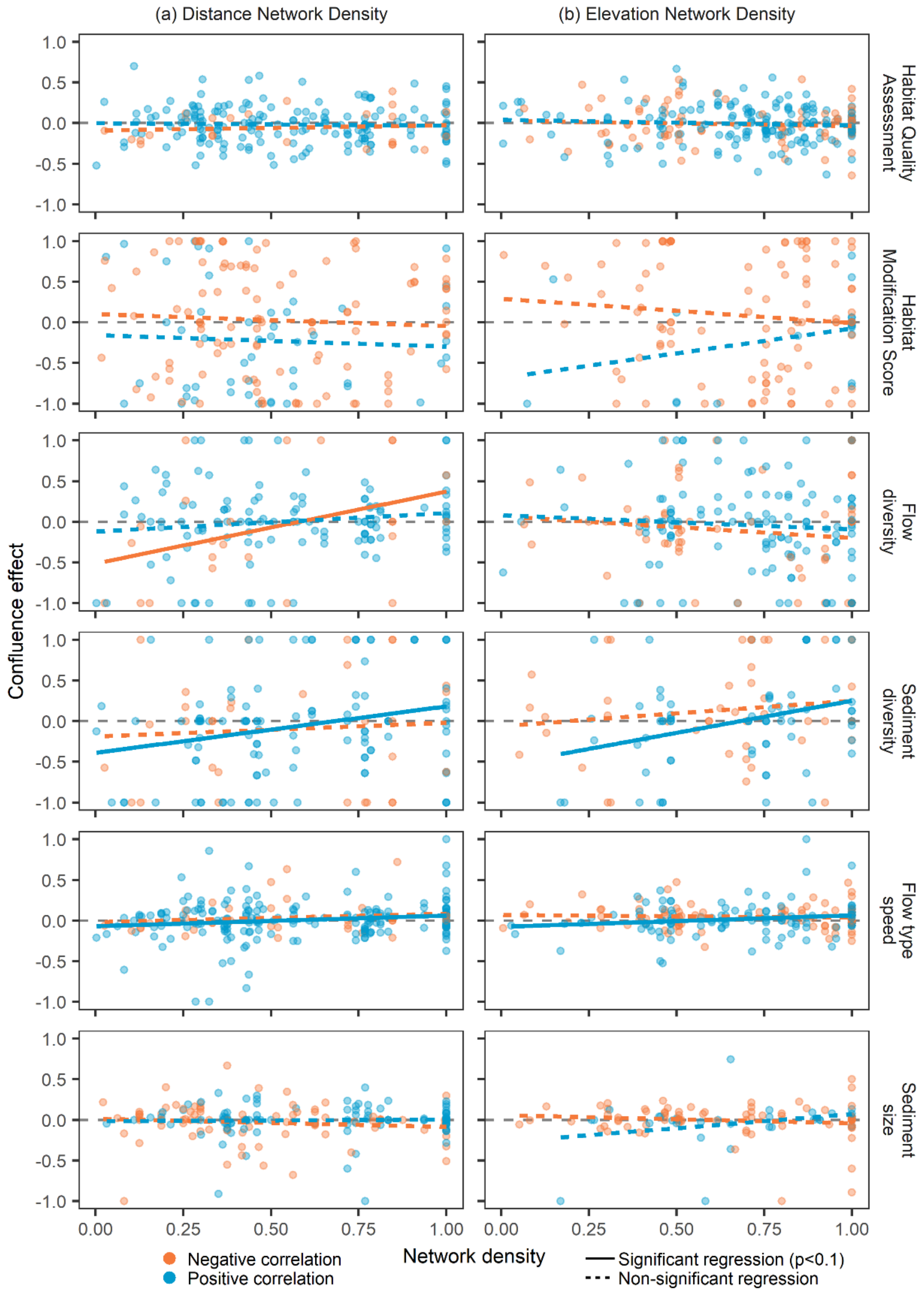


Figure 5.10. *Plots of confluence effect for each habitat index versus network density for network density metrics: (a) distance network density; and (b) elevation network density. Only catchments with significant ($p<0.05$) correlations between network density metrics and habitat indices are shown and direction of the correlation is indicated by colour. Regression lines between network density and confluence effect are plotted in bold for those relationships that are statistically different from zero ($p<0.1$).*

In contrast to the expected findings, there are often more positive confluence effects at high distance network densities in negative catchments, although only flow diversity shows a comparatively strong regression ($p=0.05$, $R^2=0.16$) (**Figure 5.10a**). Similar positive confluence effects for sediment diversity, and flow type speed are shown in high elevation network density areas in negative catchments (**Figure 5.10b**). This result implies that individual confluences may only have a positive effect on habitats within the intermediate locality (0-10^2km) and that high densities of confluences are not changing habitat patterns more widely (0-10^4km) (Rice *et al.*, 2001). High densities of confluences instead increase river discharge within a relatively short distance. This may exacerbate downstream decreases in habitat condition (Jones and Schmidt, 2016) as increases in discharge are shown to influence physical habitat type and reduce diversity (Padmore, 1998; Zavadil *et al.*, 2012).

The results indicate that in highly dense areas of the network, the effect of individual confluences can influence the impact of network topology on broad patterns of river habitats. However, in other cases, the effect of individual confluences contradicts the impact of total topological structure indicating that the effect of these confluences is too localised to influence the impact of network topology on broad patterns of river habitats.

5.4 TRIBUTARY (DIS)SIMILARITY IMPACT ON CONFLUENCE IMPORTANCE

This section presents the results and discussion for Objective 5b: to investigate which properties of upstream tributaries influence confluence importance. Previous research shows that tributaries with similar discharges create important confluences that have a positive effect on habitat features such as sediment size, geomorphic and biotic diversity in downstream reaches (see Section 5.1.1). However, there is also evidence to suggest that tributaries with dissimilar catchments create important confluences, although the effect of the confluence on downstream habitats may not be positive (see Section 5.1.2). Correlations between confluence importance and tributary properties are shown in **Table 5.2**. All 853 confluences are included in this analysis, including those in catchments with significant correlations between linear and network density metrics.

5.4.1 Strength of correlations

Correlations between confluence effect and dominant tributary properties, and between confluence strength and relative tributary properties (**Figure 5.6**) were relatively weak (**Table 5.2**). This is likely because other factors influencing confluence importance are not accounted for, such as position in the catchment, local pressures and natural variation that can confound interpretation. A similar conclusion was drawn by Jones and Schmidt (2018).

The correlations conducted also assumed a linear relationship between confluence importance and tributary properties, but the relationships may be more complex. Most studies consider confluence importance as binary (i.e. a confluence is either important or not important) (Benda *et al.*, 2004b; Jones and Schmidt, 2016; Rice, 1998, 2017) whereas this work considers confluence importance on a gradient. Jones and Schmidt (2018) also considered confluence importance on a gradient (in their case, the difference between benthic communities upstream and downstream of a confluence) and found a significant correlation between community difference and relative size of the tributaries. However, they note that the correlation is driven by one reach with differing physical conditions, and upon removal, the correlation becomes weak and insignificant. This indicates that a linear relationship is difficult to determine between tributary properties and confluence importance.

Tributary properties may also interact to influence confluence habitats, for example, tributary properties become averaged in tributaries with larger catchment areas (Fröhlich *et al.*, 2008) so larger tributaries will have similar characteristics (Jones and Schmidt, 2016). As the correlations identified with individual properties were weak, interactions are not explored in this chapter. However, interactions with relative tributary properties are investigated when multiple levels of catchment effects are combined in Chapter 6. Although the correlations were weak, there are still consistent patterns that may indicate how habitats downstream of confluences are influenced by similarities or dissimilarities in tributary properties.

5.4.2 Associations between tributary properties and confluence importance

For confluence strength, habitat condition improved consistently across all habitat indices in response to the dissimilarity of each tributary property (**Table 5.2b**) whereas for confluence effect, correlation direction differs between habitat indices for land cover and geological tributary properties (**Table 5.2a**). It must be noted that higher HMS values indicate greater modification so frequently had the opposite correlation direction to the other habitat indices.

When tributary magnitudes were similar (i.e. relative tributary property is closer to zero), confluence strength was greater (i.e. τ is negative). Also, flow type speed and sediment size increased downstream of equally sized confluences compared to upstream (i.e. τ is negative) (**Table 5.2a**). This supports previous work that found greater geomorphic diversity, maximum velocity, flow diversity and change in benthic invertebrate communities because of the creation of unique habitats downstream of equally sized confluences (Benda *et al.*, 2004a; Best, 1987; Jones and Schmidt, 2018).

Mean upstream slope was found to have stronger correlations than local slope, showing increases confluence strength with greater slope dissimilarity across all habitat indices, except for HMS which showed negative correlations (**Table 5.2b**). Improvements in habitat condition downstream were evident when the secondary tributary had a steeper upstream slope than the primary tributary (i.e. τ is negative), particularly for HQA and sediment size (**Table 5.2a**). This shows that dissimilarity in slope between tributaries is associated with confluence importance. Previous research found similar results and attributes the increase in sediment size downstream of confluences to the enhanced transport capacity of the high gradient tributary to transport coarse sediments to the confluence (Debnath *et al.*, 2019; Rice, 1998). This is important as small steep tributaries with coarse grainsizes that merge with large, low gradient main channels are common in nature and are found to create marked bed discordance that influences flow structures in downstream reaches (Leite Ribeiro *et al.*, 2012a).

Dissimilarity of selected geologies between tributaries was associated with confluence strength. In tributaries with dissimilar amounts of harder geologies, such as hard rock and limestone, the strength of confluence influence was weaker (i.e. τ is negative), whereas dissimilarities in softer geologies created stronger confluence influence on habitats (**Table 5.2b**). There was no clear evidence from the confluence effect correlations to suggest whether dissimilarity was positively or negatively associated with habitat condition (**Table 5.2a**). The exception was HMS, where modification increased downstream when the primary tributary had more hard rock than the secondary tributary.

The strongest correlations identified were with land cover dissimilarity, suggesting that wider catchment properties are related to confluence importance rather than solely the hydromorphological similarity of tributaries. Others findings support this work, finding that channel heterogeneity is more sensitive to the similarity of bedload flux and bedload grainsize between tributaries and less sensitive to discharge similarity (Rice *et al.*, 2006). This is important as bedload flux and bedload size may be influenced by the intensification of land cover (Liébault *et al.*, 2005).

Table 5.2. *Matrix of correlation coefficients (Kendall's τ) between tributary properties and the (a) confluence effect and (b) confluence strength impact on habitat indices. Colours refer to τ. *Indicates significant correlations (p<0.05)*

	(a) Confluence effect vs. dominance tributary property						(b) Confluence strength vs. relative tributary property					
Tributary property	**HMS**	**HQA**	**Flow diversity**	**Flow type speed**	**Sediment diversity**	**Sediment size**	**HMS**	**HQA**	**Flow diversity**	**Flow type speed**	**Sediment diversity**	**Sediment size**
Hydromorphology												
Stream magnitude	0.03	0.02	0.02	**-0.06***	0.00	**-0.07***	-0.01	-0.02	-0.03	**-0.10***	0.00	-0.01
Local slope	0.03	0.02	0.03	0.01	0.01	0.00	-0.03	**0.05***	0.03	0.01	0.01	0.05
Mean upstream slope	0.01	**-0.05***	-0.01	-0.03	-0.04	**-0.05***	-0.04	0.02	0.04	0.02	0.03	**0.06***
Geology												
Hard rock	**0.06***	0.02	-0.03	0.00	0.05	-0.01	0.05	-0.05	**-0.06***	-0.04	-0.05	-0.04
Other limestone	0.00	-0.01	0.01	-0.02	-0.01	-0.03	0.04	0.00	-0.02	-0.03	-0.01	-0.04
Chalk	0.00	0.00	-0.01	0.00	-0.03	-0.02	0.01	**0.07***	0.02	0.00	0.02	0.04
Sandstone	0.04	0.01	0.04	0.01	0.01	-0.01	-0.02	0.04	0.03	0.00	0.02	0.05
Other sedimentary	-0.05	0.00	-0.01	0.00	0.00	0.02	0.00	0.00	0.02	-0.02	0.04	0.05
Land cover												
Urban	0.01	0.03	0.00	-0.04	0.03	-0.03	**-0.06***	**0.10***	**0.07***	-0.02	0.01	0.05
Arable	-0.01	0.03	0.04	**0.05***	0.03	0.04	-0.03	0.03	**0.07***	0.00	0.04	**0.10***
Improved grassland	0.01	**-0.05***	-0.02	-0.01	-0.04	0.01	0.01	**-0.05***	0.00	-0.03	0.00	0.02
Semi-natural grassland	0.01	0.03	0.00	0.01	-0.01	0.00	0.02	-0.05	**-0.10***	-0.04	-0.05	**-0.12***
Mountain/ heath/bog	0.02	0.02	-0.01	-0.03	0.01	-0.03	0.01	**-0.06***	**-0.11***	-0.05	**-0.06**	**-0.15***
Woodland	0.03	0.00	-0.01	**-0.06***	0.03	-0.04	0.03	0.00	0.00	-0.03	0.00	0.01

Dissimilarity in anthropogenic land covers (such as urban and arable land) were associated with stronger confluence impacts, and dissimilarity in natural land covers were associated with weaker confluence impacts on habitat indices (except for HMS which showed the opposite result, consistent with the other correlation results) (**Table 5.2b**). Confluence effect correlations indicated that when the primary tributary was arable and the secondary tributary was not, habitat condition increased downstream of the confluence (i.e. τ is positive). This supports evidence that shows when an unregulated stream joins a regulated stream it can mitigate the negative effects of hydropower alterations on food-web structure and diversity (Sabo *et al.*, 2018).

Conversely the results also suggest that when the secondary tributary was arable, but the primary tributary was not, habitat condition worsened (i.e. τ is negative). This supports evidence of negative downstream effects when a poor quality tributary joins a high quality stream on freshwater mussel populations (Cooksley *et al.*, 2012). However, when the secondary tributary had more improved grassland or woodland than the primary tributary, HQA or flow type speed increased downstream (**Table 5.2a**).

While some habitat index responses to urban dissimilarity were consistent with arable dissimilarity, there were some differing responses for flow type speed and sediment size. These habitat indices declined downstream of a confluence with a primary urban tributary when the secondary tributary was not urban (although the correlations were not significant; **Table 5.2a**). This indicates that natural tributaries may not be able to negate some upstream anthropogenic pressures. Further work to identify when tributaries with contrasting anthropogenic pressures either mitigate or enhance negative effects downstream of confluences would be useful for targeting management efforts.

5.5 THE INFLUENCE OF CATCHMENT MORPHOMETRY ON NETWORK TOPOLOGY EFFECTS

This section presents the results and discussion for Objective 5c: to explore how catchment morphometry influences the effect of network topology on river habitats. The effect of network density metrics on habitat indices is suggested to be influenced by catchment morphometry particularly the circularity and elevation of the catchment (see Section 5.1.3). The results showed that Kendall τ values from correlations between network density metrics and habitat indices vary with catchment circularity and elevation (**Figure 5.11**).

τ tended to become more positive in circular catchments for correlations between distance network density and flow type speed in particular (τ=0.40, p<0.01), but more negative between elevation network density and habitat indices. Correlations between sediment diversity and elevation network density were the exception, where most strong positive τ values were in more circular catchments (**Figure 5.11**). The distance network density results support the NDH prediction that circular catchments will have more important confluences that increase geomorphic heterogeneity throughout the network, compared to elongated catchments where important confluences may be concentrated upstream. However, this theory assumes that the relative size of the tributary defines confluence importance (**Figure 5.5**) which is not always the case, as found in Section 5.4 and in previous studies (Duncan *et al.*, 2009). Conversely, high densities of small tributaries converging with large channels may enhance the downstream linear gradient (Jones and Schmidt, 2016) creating negative correlations in elongated catchments where differently sized tributaries are more prevalent (as shown in **Figure 5.5**).

Mean elevation had a positive influence on τ values for correlations between distance network density and the sediment size index in particular (correlation between mean elevation and the τ value for the association between distance network density and sediment size; $\tau=0.38$, $p<0.01$). This suggests that habitat condition worsens with distance network density in lowland catchments but improves in upland catchments. This supports empirical studies in high-energy environments that show positive effects from important confluences (Benda *et al.*, 2004a; Rice *et al.*, 2001) because of greater coarse sediment supply and transport capacity (Ferguson *et al.*, 2006) which are features of upland catchments in the UK (Raven *et al.*, 2010). Lowland, low-energy or modified catchments in comparison have confluences that have negative or no important effect on habitats downstream (Singer, 2008). This may be because land covers that are more prevalent in lowland catchments, such as arable and urban land covers, may have negative effects on habitat indices downstream of confluences.

For correlations between habitat indices and elevation network density, there are few associations with catchment morphometry. However, catchments with lower mean elevation have stronger correlations than higher elevation catchments, although the direction of the correlations vary (**Figure 5.11**). The directional variation in correlations may be due to greater anthropogenic pressures in lowland streams causing negative effects at confluences in some catchments.

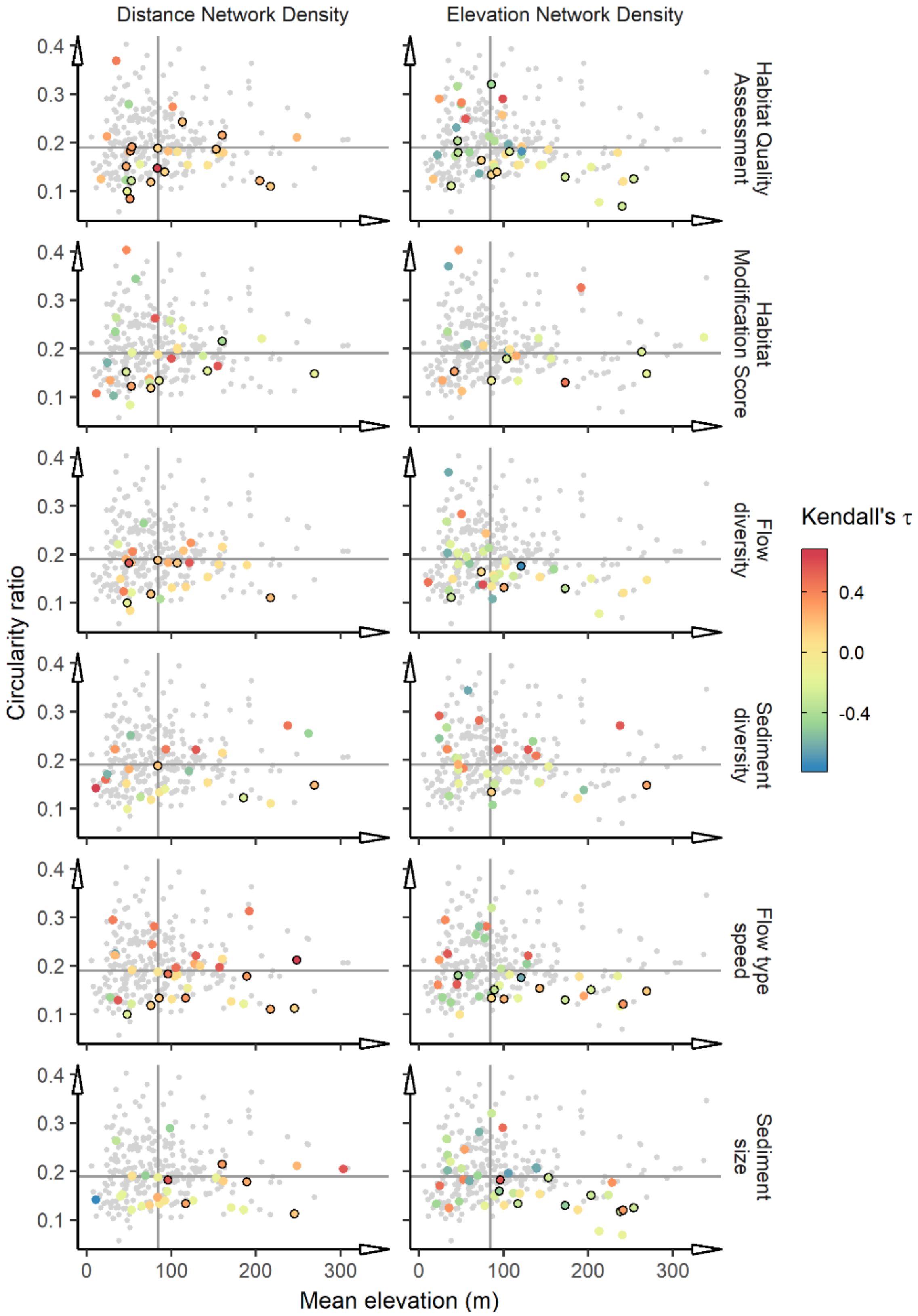


Figure 5.11. *Kendall's τ of catchments with significant correlations (p<0.05) between network density metrics (columns) and habitat indices (rows) plotted on mean elevation and circularity axes. Bold points are significant (p<0.05) after FDR correction to the p-value. Grey points indicate catchments with no significant correlation. Lines indicate median elevation and circularity ratio values for all catchments where correlations were calculated.*

5.6 DISCUSSION AND CONCLUSIONS

Network topology influences the type and diversity of physical habitat found in river channels. It is a component of the catchment that is often not considered or poorly quantified in catchment-level studies (see evidence review in Section 2.1.3). Chapter 4 presented metrics that easily quantify network density (and as a result, confluence density) in catchments, opposed to downstream continuums of habitat types (**Figure 5.1**).

Network density metrics describe downstream patterns in habitats better than the downstream linear gradient in the four example catchments (Chapter 4). However, both linear and network density metrics showed few significant correlations with habitat indices within 348 networks in England (**Figure 5.7**). This is logical as rivers in Britain have been modified since the Bronze Age (Macklin and Lewin, 2003) so external factors not considered in this analysis will likely have a strong control on river reach function. It would be interesting to repeat this studied in an un-modified landscape, although broad-scale habitat data would likely unavailable. Despite this, network topology is still an influential control on physical habitats in around 10% of networks, but areas of high network density did not always generate improved habitat conditions as anticipated (**Figure 5.7**). This chapter explores the causes for this in the guise of three objectives that stem from the idea that not all confluences have an important effect on the channel (Rice, 1998) and that effects may not always be positive. This was an assumption made in Chapter 4 to test the hypothesis that highly network dense areas in catchments yield better habitat conditions. The results of the objectives, detailed in Sections 5.3 to 5.5, are summarised in **Figure 5.12** and discussed in wider contexts below.

Statistically, habitat condition improves downstream of confluences (**Figure 5.8**) but not all confluences have a positive effect on habitats (**Figure 5.9**). Results indicate that when important confluences that improve habitat condition are in located in high density areas, there are positive associations between network density and overall habitat condition (supporting the ideas of the NDH; Benda *et al.*, 2004b). However, this may be overridden when important confluences that have negative effects are located in high density areas, causing negative correlations with habitat condition. Confluence importance therefore influences the association between habitat condition and network topology rather than solely the density of the network. There is also evidence that the relationship between habitat condition and network density may not reflect confluence effect for some habitat indices. This signifies that confluences may change habitat condition locally, but not influence habitat condition in the wider area (Rice *et al.*, 2001).

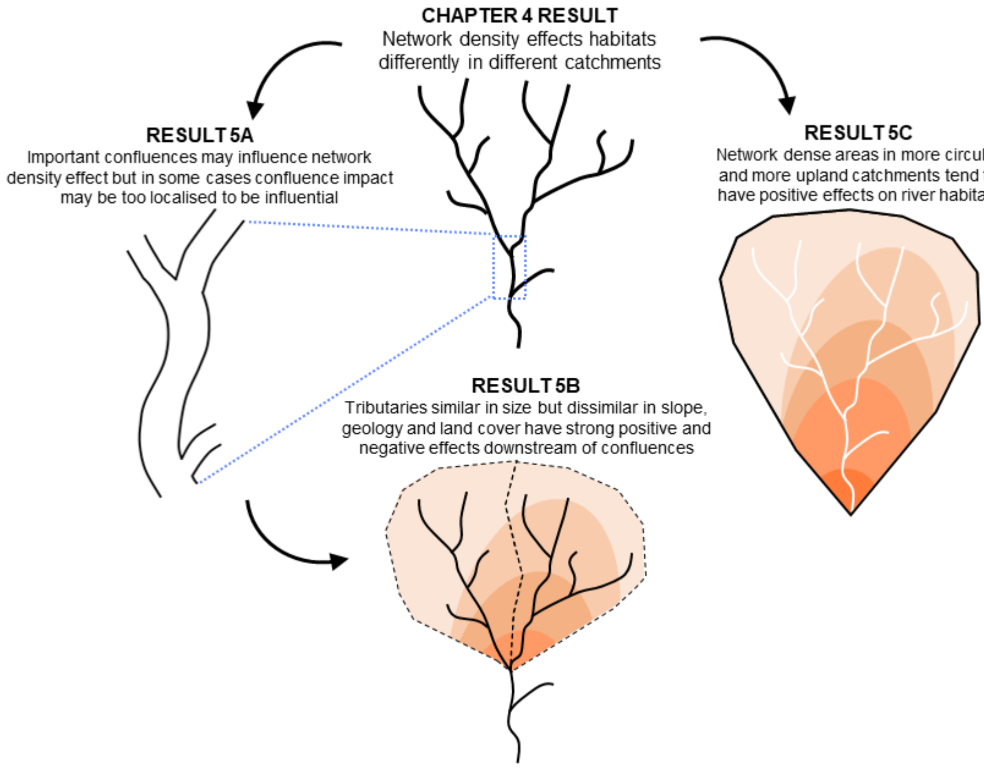


Figure 5.12. *Rework of* ***Figure 5.2*** *outlining the key results to each objective.*

Tributaries of a similar size and/or dissimilar properties are often the attributed controls on confluence importance, but most previous work has been conducted in selected high-energy, natural catchments, rather than numerous modified catchments as used in this study. While correlations between tributary dissimilarity and confluence effect are weak, there are clear patterns (**Table 5.2**). Equally sized tributaries create confluences with a stronger effect on habitat features, with faster flow types and coarser sediments present downstream which supports previous findings (Benda *et al.*, 2004a; Best, 1987; Jones and Schmidt, 2018). Tributary dissimilarity is also important with poor-quality tributaries worsening habitats and high-quality tributaries potentially mitigating poor habitat conditions of the primary channel. However, natural tributaries are not always found to be effective at mitigating upstream negative influences on habitats. This is one of the largest empirical studies on confluence effects (see Section 5.3.2) and demonstrates that relationships between tributary properties and confluence effects are complex and not always positive, a finding that has thus far not been included in the key studies in this field (e.g. the NDH; Benda *et al.*, 2004b). Further work to successfully predict which confluences are most important for altering reach-level features, by utilising statistical methods that could accommodate non-linear relationships and interactions with other variables would be the logical next step. A useful development of this work would be to calculate network density metrics for only important confluences to see if this had a greater influence on habitat than the network topology metrics of the entire network.

Catchment morphometry was also found to influence the effect of network topology on physical habitats. Positive associations between network density and habitat condition are found in more circular catchments (i.e. catchments with assumed more important confluences, supporting the NDH; Benda *et al.*, 2004b; **Figure 5.5**) and negative relationships in elongated catchments, where many non-important tributaries accelerate downstream declines in habitat condition (**Figure 5.11**). This supports previous findings from this work (Section 5.4) and others (Ferguson *et al.*, 2006; Rice *et al.*, 2006) that show that confluence effect differs with the size similarity of the tributaries and properties that impact sediment supply and transport capacity, altering physical habitats. There are also more negative correlations in lowland catchments, where tributaries are less natural and have less sediment transport capacity.

The statistical relationships identified with network topology in this chapter are relatively weak, and reasons for this are discussed in Sections 5.3 to 5.5. The reliance on multiple catchment-level correlations makes it difficult to identify specific areas where network topology is important. Future research should focus on individual catchments, such as the work in Chapter 4, in order to better understand the implications of network topology. This work also treats both the network and confluences features as static when processes such as seasonal network expansion of ephemeral streams (Wharton, 1994) and confluence mobility over longer timescale (Dixon *et al.*, 2018) will alter the longevity of these results.

In conclusion, although the landscape of England is highly modified, theories of network topology developed in high-energy natural systems are still relevant and should be considered by future catchment-level studies in low-energy and/or modified landscapes. The nature of such landscapes means that confluences, and high densities of them, may produce both better, poorer and most likely no difference in habitat condition compared to surrounding areas. However, there is evidence that higher energy, natural tributaries and catchments produce more positive habitat effects in response to confluences and high network densities. The results here and the results of others (Kiffney *et al.*, 2006; Leite Ribeiro *et al.*, 2012b) indicate that management plans should include the entire drainage system and target tributaries that cause negative effects downstream whilst conserving tributaries that mitigate negative effects from upstream.

Understanding river habitat response to catchment-level effects at multiple spatial levels

6.1 INTRODUCTION

Each previous data chapter (Chapters 3 to 5) has quantified a property of the catchment-level effect at a different spatial level in the river system hierarchy and explored the impact on physical habitats type and diversity within reaches. This chapter presents an analysis of the outputs of previous data chapters to form a synthesis to the thesis that incorporates multiple spatial levels and influences (**Figure 6.1**).

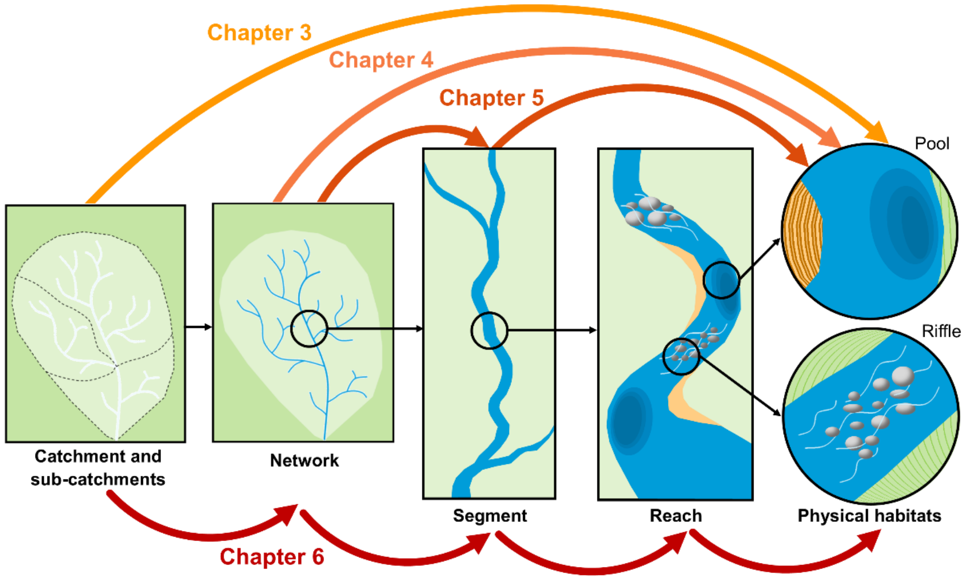


Figure 6.1. *Reprint of* ***Figure 1.2.*** *Multiple spatial levels explored in the thesis in each chapter and how this chapter (Chapter 6) combines the multiple outputs.*

The hierarchical nature of the river system is a critical conceptualisation, demonstrating how processes occurring at the catchment-level can affect processes occurring at lower levels (Section 2.1.2). Hierarchical levels have been considered in previous studies utilising the RHS dataset to explain wider influences on river habitats (see Section 6.1.1 for review) and forms a key component of some river typologies (e.g. REFORM; Gurnell *et al.*, 2016). The evidence review in Section 2.1.3 revealed that most large-scale studies of catchment-level effects focus on a few characteristics of the catchment, namely land cover, geology and the catchment area upstream of a site (**Figure 2.5**), often failing to consider components of the river network and considering individual characteristics in isolation, without accounting for their interactions. The thesis so far has produced a typology of the combined effect of multiple terrestrial catchment characteristics (Chapter 3), identified meaningful metrics of network topology (Chapter 4) and explored the relative influence of upstream tributaries on confluences (Chapter 5). The aim of this chapter is therefore to understand how these components at different levels in the hierarchy interact to explain patterns of physical habitats and thus better define the elusive catchment-level effect. This is useful for management hoping to restore or conserve habitats so that plans may be implemented strategically with an appreciation of holistic catchment-level effects. The specific objective of the chapter is as follows:

Objective 6a: To use a range of GIS-derived catchment and river properties at different spatial levels, to explain patterns of physical river habitats in England.

To address the objective, this chapter first describes which characteristics have been explored by previous studies using the RHS and which methods they have used (Section 6.1.1). Then the multi-level characteristics used by this study and the methods of data exploration and model development are described (Section 6.2) before results are presented (Section 6.3). A discussion and synthesis are provided in Section 6.4 before the chapter conclusions.

6.1.1 Effects on physical habitats from different spatial levels: Examples using the River Habitat Survey

Since the RHS's inception in 1994, multiple attempts have been made to explain patterns of physical habitats extracted from the dataset. A summary of these studies is provided in **Table 6.1**. Two key trends in research are observed and are discussed in the sections below.

6.1.1.1 Movement beyond reach-level controlling variables

The first trend observed is that many early studies focused on reach characteristics derived from GIS and sub-reach characteristics measured during RHS surveys. The seminal study from Jeffers (1998) explains patterns in individual RHS features (such as sediment types, flow types, channel features etc.) using two principal components (PCs) derived from altitude, slope, distance to source and height of source for each RHS site. The two principal components represent upland to lowland and high energy to low energy gradients (**Figure 6.2**). The components have been used since to describe distributions of RHS sites in Britain (e.g. Harvey and Wallerstein, 2009) and are included in the RHS dataset provided by the Environment Agency. These reach-level characteristics describe the position of the RHS site within the catchment which is proposed by Vannote *et al.* (1980) among others, to reflect a continuum of controls on river condition from upstream to downstream (see Section 2.1.1.1 for review of continuum theory).

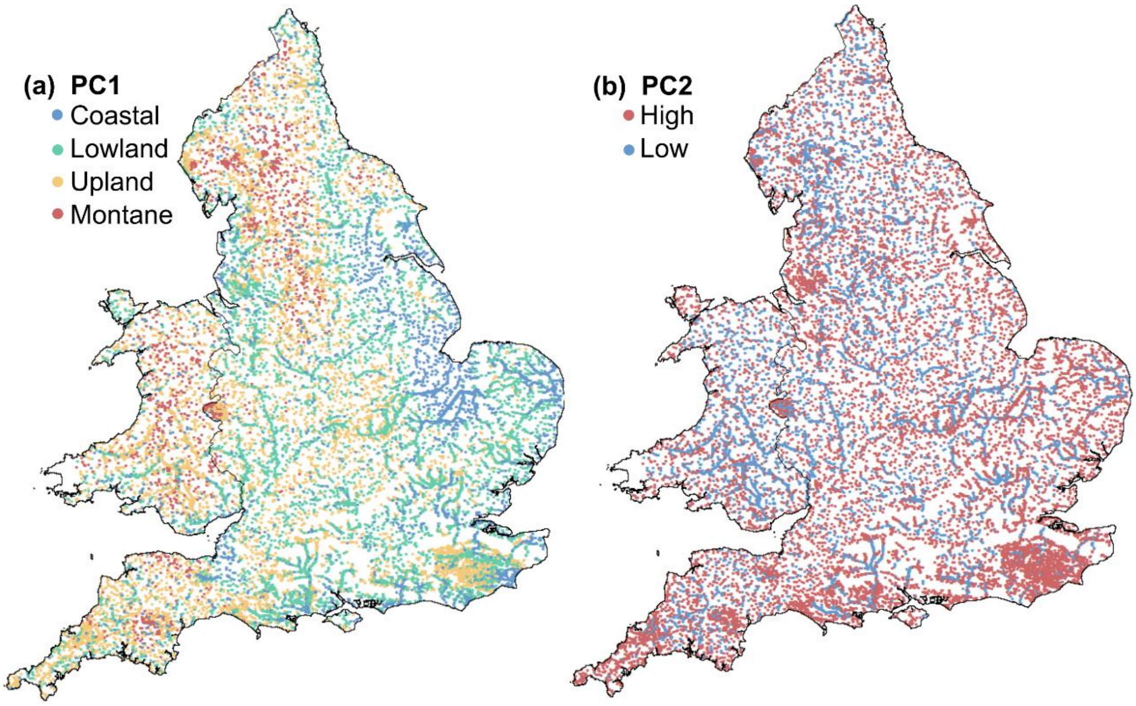


Figure 6.2. *RHS sites in England and Wales classified according to Jeffers' (1998) PC1 (upland-lowland) and PC2 (high-low energy).*

There are studies considering a wider range of factors including catchment-level characteristics but they are limited in number (Jusik *et al.*, 2015; Naura *et al.*, 2016; Vaughan *et al.*, 2013). Characteristics reflecting geology and hydrology are often found to be significant within models, potentially due to their dominant control on catchment and network formation (Schumm and Lichty, 1965). Only one study considered morphometry and rainfall parameters and frequently found them to be significant predictors of hydromorphological features (Vaughan *et al.*, 2013). Land cover was considered by only two studies and was not included in final models in either (Jusik *et al.*, 2015; Naura *et al.*, 2016). This is surprising given the results of the evidence review in Section 2.1.3 that found that land cover was most frequently included in studies of controls on river habitats.

Vaughan *et al.* (2013) compared the predictive capabilities of their multi-level model (**Table 6.1**: Model 1) to a reduced model containing only stream power and upstream area (**Table 6.1**: Model 2) and to Jeffers' (1998) PCs (**Figure 6.2**). The multi-level model is based on a regression of the three PCs: PC1, representing an energy gradient; PC2, an upstream-downstream gradient and; PC3, groundwater contribution. They found that the stream power-area model reflected similar gradients to the multi-level model and accounted for almost as much variation in habitat features. Both models were superior to Jeffers' principal components (**Figure 6.2**). They therefore suggest that catchment area and stream power, which reflect Jeffers' upland-lowland and energy gradients respectively, may be the dominant controls on river features. However, while statistically independent, catchment area and stream power are not functionally independent. This is because a component in the stream power equation (median annual discharge) is estimated in part from catchment area in Vaughan *et al.*'s (2013) study, limiting the understanding gained from the reduced model. It is also difficult to understand interactions and isolate the individual controls on habitat in the multi-level model from the PCA.

The research using RHS to date suggests that physical habitats are controlled by reach and segment-level parameters, especially elevation, slope and unit stream power (**Table 6.1**). Research using catchment-level parameters is limited and, for the most part, considers only catchment characteristics, such as geology, at the RHS site rather than geological properties upstream influences which are of known importance (Section 2.1.3). As far as the literature search here can identify, none of the studies thus far have considered network topology beyond that of stream order (for discussion around limitations of stream order see Section 4.2.1.1) or the influence of confluences.

6.1.1.2 Changing objectives over time

The second trend observed is a shift from explanatory to predictive statistical approaches reflecting and overall trend from theory-based to data-centric science (Karpatne *et al.*, 2017). The objective to produce an accurate predictive model of physical habitats stems from the inability to sample all reaches due to limitations in time and resources. By effectively predicting habitat features, the need for expensive, time consuming surveys is reduced, and inferences can be made based on GIS-derived data. This was the approach taken by Naura *et al.* (2016) whereby sediment size is inferred through regression kriging in un-surveyed reaches. The production of effective predictive models, however, often comes at the expense of understanding the mechanisms and controlling factors on the response variable. For example, often PCs are used to reflect multiple control and/or response variables in the regression models as a method of data reduction and to improve predictive accuracy (**Table 6.1**). However, this makes it difficult to isolate which responses are being influenced by which parameters.

There is also little consideration in previous studies of the hierarchical structure of the multi-level data and the emergent effects in multivariate analysis when parameters are either combined into a few PCs (Vaughan *et al.*, 2013) or entered as individual parameters in predictive models. Statistical approaches such as clustering and tree based approaches adopted by some studies (Bizzi and Lerner, 2015; Harvey *et al.*, 2008a) can show the hierarchical structure of these data and interactions between levels, however, these studies have focused on the sub-reach to segment-levels thus far (**Table 6.1**).

Table 6.1. Variables included in previous attempts to predict RHS data. X variable has stronger effect in model, + variable considered but either weak effect or removed from final model.

			Sub-reach		**Reach**							**Segment**				**Catchment**																	
																Morphometry				Rain		Geology								Hydrology			
Study	**Method**	**Response**	Other habitat features	Modification scores	Altitude	Slope	Distance to source	Height of source	Unit stream power	Catchment area	Stream order	Floodplain confinement	Discharge estimate	Total stream power	Stream power change	Mean slope	Elevation-relief ratio	Std dev elevation	Mean aspect	Mean annual rainfall	Days >10mm rain	Hard rock	Sandstone	Other sedimentary	Limestone	Chalk	Solid geology	Drift geology	Solid geology age	Major aquifers	Hydrometric areas	Base flow index	Land cover
Jeffers, 1998	PCA	Individual features			X	X	X	X																									
Harvey *et al.*, 2008a	Hierarchical clustering	Individual features	X																														
Harvey *et al.*, 2008b	Kruskal-Wallis	PCA feature scores																				X	X	X	X	X							
Harvey and Wallerstein, 2009	Non-statistical	Individual features		X																													
Bizzi and Lerner, 2012	Self-organising clustering	Morphology classes				X			X		X	X	X																				
Vaughan *et al.*, 2013– Model 1	Generalised Additive Model	Individual features			X*	X*	X*	X*	X*	X*		X*				X*	+*	X*	+*	X*	X*	+*	+*	X*	+*	X*					X*	X*	
Vaughan *et al.*, 2013– Model 2	Generalised Additive Model	Individual features							X	X																							
Bizzi and Lerner, 2015	Classification tree	Morphology classes							X			X		X	X																		
Jusik *et al.*, 2015	Monte Carlo permutation	Macrophyte clusters	X	+	X	+				+																							+
Naura *et al.*, 2016	Regression kriging	CA sediment type			X*	X*	X*	X*																			X	+	X		X		+

* Regression conducted on PCA

Variable calculated as % within upstream catchment, all other variables extracted at site

Often parameters that successfully predict a response variable are said to influence the response variable, but there is an argument that a single 'best' predictive model may have little explanatory power if it is not based in scientific theory. This is due to multicollinearity, where successful predictors may imitate the interaction of variables that cause the variation in the response variable.

Therefore, in order to explain the response, rather than predict it, multiple variables that should explain patterns of physical habitats according to scientific theory (such as the results in this thesis and from others) must be considered and combined using data-science techniques (Mac Nally, 2000). This explanatory approach has been adopted in studies exploring the effects of catchment morphology on channel features (e.g. Jensen *et al.*, 2018). Thus, the success of statistical models is assessed not on predictive power but on the ability to offer mechanistic explanations of variable interactions.

6.2 METHODS

There are challenges when conducting multivariate analysis on broad-scale monitoring data (Feld *et al.*, 2016). Challenges include the high number of control variables, many of which exhibit collinearity, non-parametric skewed distributions dominated by zeros and non-linear relationships with response variables. Feld *et al.* (2016) makes recommendations of how to address these challenges, specifically in relation to aquatic monitoring data, which are adopted here.

First, according to Feld *et al.* (2016), key gradients in these data must be explored which is conducted here with Principal Component Analysis (PCA). Second, regression trees are built for each of the four habitat indices extracted from the RHS dataset (**Table 2.2b**), to explore the hierarchical structure of relationships of the multi-level variables to physical habitats. Finally, this analysis is extended using the machine learning technique of 'boosting' to produce boosted regression tree (BRT) models for identifying the most important controls variables contributing to the catchment-level effect on physical habitats, and their interactions.

6.2.1 Multi-level GIS-derived variables

Control variables are chosen to reflect known drivers of channel form and function that are commonly included in large-scale studies (e.g. reach and upstream variables, see Section 2.1.3). Also included are variables that have not before been included in large-scale multivariate studies (e.g. network and confluence variables) that can easily be derived from spatial data in a GIS. This totals 35 individual GIS-derived variables representing the multiple levels in catchment-level effects (**Table 6.2**). Only GIS-derived variables are selected so that relationships may be extended to un-surveyed reaches.

Upstream-level variables reflect external drivers on river habitats (see evidence review in Section 2.1.3) and are calculated here as the percentage of land cover or geology within a 100m buffer of the network upstream of the RHS site. Characteristics are calculated for a buffer rather than within the entire upstream catchment boundary due to ease and accuracy of computation (for further discussion see **Appendix 5C**).

Network-level variables reflect the topological signal of the entire network. Network density is calculated for distance or elevation bands within each catchment and normalised between zero and one (with one being the highest density) within each catchment. Network density is shown to reflect up-to-downstream changes in physical habitats in some catchments in England (Chapter 4 and Section 5.3).

The sub-catchment-level variable, waterbody type, reflects 22 GIS-derived terrestrial characteristics including geology, land cover, climate and morphometry to represent the combined drivers' effects within the waterbody local to the RHS site. The waterbody types have been shown to reflect broad patterns in average habitat type and diversity in England (Chapter 3).

Confluence variables at the segment-level reflect the relative contribution of the two incoming tributaries upstream of each RHS site. The relative difference in hydromorphology characteristics (area, magnitude, slope and stream power) between upstream tributaries has been shown by this thesis (Section 5.4) and previous studies (Benda *et al.*, 2004a; Best, 1987; Jones and Schmidt, 2018) to influence physical habitats. The relative amount of upstream land cover and geology has not before been considered by many previous studies as an influence on physical habitat but differences between tributary properties is suggested to increase the effect on physical habitats at confluences (Section 5.4).

Reach-level variables are included that describe the position of an RHS site within the region, catchment and network and have been suggested to reflect upland-lowland and energy gradients in the UK by Jeffers (1998a) and others (Section 6.1.1).

Table 6.2. *GIS-derived variables at multiple levels used in the statistical analysis, with abbreviations and descriptions. Where a variable is calculated in a previous chapter, the chapter is indicated in square brackets. For data sources see Section 3.2.2.1.*

Level	Variable (units)	Abbreviation	Description
Upstream	<u>Geology:</u>		Upstream percentage cover of geological of land cover characteristic within a 100m buffer of network upstream of RHS site.
	Hard rock (%)	Ups.Hard	
	Other limestone (%)	Ups.Lime	
	Chalk (%)	Ups.Chalk	
	Sandstone (%)	Ups.Sand	
	Other sedimentary (%)	Ups.Sed	
	<u>Land cover:</u>		
	Woodland (%)	Ups.Wood	
	Mountain/bog/heath (%)	Ups.Mount	
	Seminatural grassland (%)	Ups.NatGrass	
	Improved grassland (%)	Ups.ImpGrass	
	Arable (%)	Ups.Arable	
	Urban (%)	Ups.Urban	
Network	Distance network density (0-1) **[Chapter 4]**	Net.D.Density	Density per distance band, normalised per catchment
	Elevation network density (0-1) **[Chapter 4]**	Net.E.Density	Density per elevation band, normalised per catchment
Sub-catchment	Waterbody type (categorical) **[Chapter 3]**	Waterbody	7 categories:
		UG	*Upland grassland*
		UN	*Upland non-grassland*
		SL	*Seasonal*
		MR	*Mid-range*
		AQ	*Aquifer*
		LA	*Lowland arable*
		LU	*Large urban*
Confluence	**[Chapter 5]**		
	<u>Upstream geology:</u>	Rel.Hard [a]	Relative difference in percentage cover between the primary and secondary tributary upstream of RHS site. Calculated as:
	Hard rock (%)	Rel.Lime [a]	[a] Total difference
	Other limestone (%)	Rel.Chalk [a]	[b] Relative difference
	Chalk (%)	Rel.Sand [a]	
	Sandstone (%)	Rel.Sed [a]	
	Other sedimentary (%)		
	<u>Upstream land cover:</u>		
	Woodland (%)	Rel.Wood [a]	
	Mountain/bog/heath (%)	Rel.Mount [a]	
	Seminatural grassland (%)	Rel.NatGrass [a]	
	Improved grassland (%)	Rel.ImpGrass [a]	
	Arable (%)	Rel.Arable [a]	
	Urban (%)	Rel.Urban [a]	
	<u>At confluence hydromorphology:</u>		
	Slope (degrees)	Rel.Slope [b]	
	Cumulative catchment area (km^2)	Rel.Area [b]	
	Shreve order/magnitude (variable units)	Rel.Shreve [a]	
	Total stream power (variable units)	Rel.Power [b]	
Reach	Cumulative catchment area (km^2)	Rch.Area	Values extracted at RHS site
	Slope (degrees)	Rch.Slope	
	Strahler stream order (variable units)	Rch.Order	
	Elevation (m)	Rch.Elevation	
	Distance from mouth (km)	Rch.Distance	
	Distance of site to upstream confluence (km)	Rch.Confluence	

6.2.2 Data-science analysis

6.2.2.1 Explorative ordination

The control variables are explored with Principal Component Analysis (PCA) to identify key gradients and correlations within these data before multivariate analysis is conducted (Feld *et al.*, 2016). Data reduction has already been applied to the sub-catchment-level waterbody type variable to produce seven types (Section 3.2) which are mapped onto the PCA biplots, along with the habitat indices, to identify structure within the control variables. Two separate PCAs are also created for the reach-level variables only and for the non-reach-level variables only to identify factors driving gradients in the data.

6.2.2.2 Regression trees

Regression trees are appropriate for addressing the issues with the multi-level data as they accommodate different types of variable, missing data, nonlinear relationships, are insensitive to outliers, automatically reject irrelevant variables and account for variable interactions. Regression trees work by splitting the parameter space into rectangular regions that have a homogeneous relationship between the response variable and control variables. This process often produces very large trees from recursive binary splitting and is very sensitive to training data (Hastie *et al.*, 2001). While this causes uncertainty and limited predictive performance, single regression trees are an intuitive method of visualising data partitions (Elith *et al.*, 2008).

Here, regression trees are constructed for each habitat index response variable with the '*rpart*' package v4.1.13 in R (Therneau and Atkinson, 1997) on training data (75% of total dataset) that is created by grouping RHS surveys into percentiles based on the habitat index in question and randomly sampling within these percentiles to retain data structure in the training data using the '*caret*' v.6.0.84 package (Kuhn *et al.*, 2019). The predictive capability of the trees is validated against the remaining 25% of the dataset as a measure of confidence in the results. More levels in the tree produces less classification error on training data but may cause overfitting (Hastie *et al.*, 2001) so trees are pruned to prevent overfitting (two methods of pruning are compared in **Appendix 6A**).

6.2.2.3 Boosted regression trees

Machine learning approaches are recommended to deal with inherent issues with analysing multivariate data (Feld *et al.*, 2016). Boosting is a machine learning method that has been applied to regression trees by stochastically selecting subsets of the dataset to build each tree (**Figure 6.3**). This improves model accuracy by averaging many rough models rather than trying to fit one highly accurate model. Boosted regression trees (BRTs) are shown to give superior results compared to linear models when variables are highly skewed, which is often the case with environmental datasets (Elith *et al.*, 2008).

BRTs are appropriate for these data as they require a large number of observations to produce a stable result (Feld *et al.*, 2016) which is available with the RHS dataset, whilst also retaining the advantages of tree based methods described in Section 6.2.2.2. They have been used in river research with the aim of predicting characteristics such as stream biotic communities (Waite *et al.*, 2019) and eutrophication levels (Rankinen *et al.*, 2019) from external controls. BRTs are not only a predictive tool but identify the most influential variables so can be used for explanatory studies (Perry *et al.*, 2012).

BRTs work by iteratively fitting decision trees to training data using cross-validation (CV) rather than single training and validation datasets (as used for the single regression trees in Section 6.2.2.2) so that all these data can influence the final model (**Figure 6.3a**). The first tree fitted achieves the best possible reduction in the loss function for a single tree. Subsequent trees are fitted to the residuals of the previous tree, putting the emphasis on fitting the poorly modelled observations. The stagewise process means that the original model is maintained.

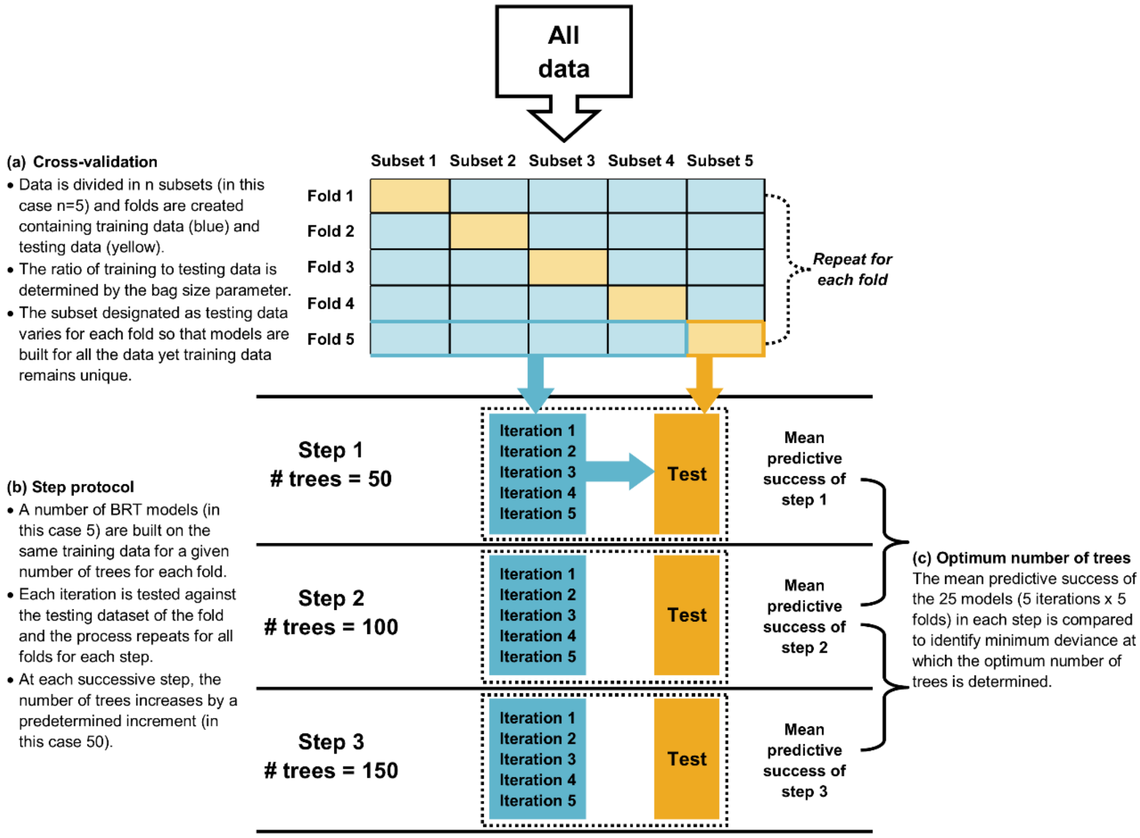


Figure 6.3. *Schematic of Elith et al.'s (2008) cross-validation (CV) protocol to determine the optimum number of trees in three steps: (a) cross-validation; (b) step protocol; and (c) determining the optimum number of trees.*

Model tuning
The number of trees built in the BRT is defined by the user so the optimal number of trees must be identified for each BRT. The optimal number of trees varies based on two tuning parameters: learning rate (the contribution of each tree to the growing model) and tree complexity (the number of interactions between variables accounted for). The tuning parameters are systematically varied (**Appendix 6B**) to identify the optimum number of trees using the CV '*gbm.step*' subroutine implemented in the '*dismo*' v.1.1.4 add-on for ecological data (Hijmans *et al.*, 2017) to the '*gbm*' v.2.1.5 package (Brandon *et al.*, 2019) in R (**Figure 6.3**). The optimum number of trees is identified at minimum predictive deviance if there is: (i) more than 1000 trees, according to Elith *et al.*'s (2008) rule of thumb, and; (ii) no evidence of overfitting (steep increases in deviance after reaching minimum). This CV approach is increasingly used for selecting optimal model settings (Hastie *et al.*, 2001).

Using this method, the selected BRT tuning parameters are learning rate = 0.01 and tree complexity = 7. Other parameters are set at the defaults recommended by Elith *et al.* (2008) and as employed by Messina *et al.* (2019): bag size (percentage of data given to the training dataset for cross validation) = 0.75; number of folds = 10 and; step size (number of trees built at each step) = 50. The optimum number of trees varies for each BRT model built to explain each habitat index response variable. The code for producing the BRT models is provided in **Appendix 6C**.

Assessing model performance
The performance of the four BRT models, one for each habitat index, is assessed by the proportion of variance explained by the model using R^2 values (Moriasi *et al.*, 2007). Performance can be calculated by testing the predicted against the observed values, but this can lead to inflated measurements (Waite *et al.*, 2019). A more conservative approach is to evaluate the models against data withheld from their calibration in the CV process, the 25% bag-fraction (Waite *et al.*, 2019). As the primary goal of this analysis is explanatory rather than predictive, quantitative measures of performance are not the focus, but the performance measures do allow for the assessment of confidence in the models for explaining habitat indices (Section 6.3.3.1).

It is not recommended to view individual trees within the BRT models (Brandon *et al.*, 2019) which means it is harder to interpret the hierarchical interactions of the predictors than in a single regression tree. However, the relative importance of the variables (Section 6.3.3.2) and the interactions between predictors (Section 6.3.3.3) may be assessed.

The relative importance of variables to the model is calculated based on Friedman's (2001) formula accounting for the number of times a variable is selected for splitting and the improvement to the model as a result of each split. Partial dependence plots are used to show the effect of control variables, both individually and when interacting with other variables, on the habitat indices while other predictors in the model are held at their mean value (Friedman, 2001). The fitted functions in the partial dependence plots may not reflect the distribution of observations so should be interrogated for step changes and general trends, rather than minor fluctuations. The strength of interaction between variables in the model is estimated while all other variables are set to their respective mean values, as described by Elith *et al.* (2008), with higher values indicating stronger interaction effects.

6.3 RESULTS

6.3.1 Principal Component Analysis

This analysis is an exploratory analysis step, so the data structure and key axes of the datasets are the focus rather than the percentage variance explained and specific loading values (which are provided in **Appendix 6D**).

The first three principal components (PCs) of the PCA model containing variables from all hierarchical levels (**Figure 6.4a**) reflect those identified by Jeffers (1998) and Vaughan *et al.* (2013) and is supplemented by the additional confluence and network variables included in this analysis.

- PC1 reflects the upland-lowland gradient from high, steep reaches with hard-rock, limestone, mountainous land or natural grassland to reaches in lower, flatter landscapes with arable or urban land covers.
- PC2 reflects the size and location of the reach with small values indicating downstream reaches with large catchment areas to reaches further from the mouth. The confluence and network variables respond to this axis with greater differences between upstream tributaries in downstream reaches whereas upstream reaches are more homogeneous but have greater network density.
- PC3 is a gradient from reaches with groundwater inputs upstream (sandstone geology) and a similar area, slope and power contributions from tributaries to reaches with differing area, slope and power between tributaries in catchments dominated by other sedimentary geology.

The waterbody types respond primarily to PC1 with upland, midland and lowland types ordered along the first axis. The upstream-downstream gradient (PC2) shows a mix of waterbody types along the axis but is dominated by large urban waterbody types at the downstream end of the gradient, reflecting the large catchment areas draining into these waterbodies. The aquifer waterbodies are frequently found towards the arable, sandstone and chalk arrows. Variation in waterbody type is most frequently associated with the confluence variables reflecting differences between tributaries which is likely because the confluence variables encapsulate variation within the wider catchment opposed to homogeneous sub-catchment waterbody types (**Figure 6.4b**).

When only the reach-level variables are selected for the PCA (**Figure 6.5a**), the primary axis showing the upland-lowland gradient is highly dominant with the secondary axis primarily driven by the distance to confluence variable. PC2 differentiates the aquifer waterbody type, due to the low drainage density present in groundwater fed catchments, but the lowland arable, mid-range and seasonal waterbody types are grouped together with little differentiation. When reach-level variables are removed from the PCA (**Figure 6.5b**) the distribution of RHS sites within the biplot is similar to the full model (**Figure 6.4b**). The upland-lowland gradient is still dominant, but the secondary size gradient is less pronounced due to the removal of the stream order and cumulative catchment area reach variables. PC2 is instead a gradient from reaches with highly different upstream tributaries but low network density to similar tributaries in highly network dense areas. This highlights the importance of extending multivariate analysis beyond the reach-level to capture a wider range of variation in controls on instream habitats.

Of the four habitat indices explored by this thesis, flow diversity, flow type speed and sediment size respond to the PCA biplot, showing increasing values along the upland-lowland gradient (**Figures 6.4c, e and f**). This is particularly evident for the coarsest sediment values with average calibres greater than coarse pebbles (sediment size index >5, **Table 2.3c**) predominantly present >0 on PC1.

While there are general trends, there is also variation that may be explained by the other axes. For example, reaches with average silt sediment (sediment size index <-5, **Table 2.3c**) are not present at the lowest end of the upland-lowland gradient, but better reflect the amount of upstream urban land cover or the relative amount of arable land cover (**Figure 6.4a**). Also, flow diversity shows increasing diversity along PC2 as well as PC1 (**Figures 6.4c**). This variation of habitat indices along the PCA axes highlights that attempting to understand the variability in river habitats across England with a single regression line or equation is insufficient without accounting for other structures in the dataset.

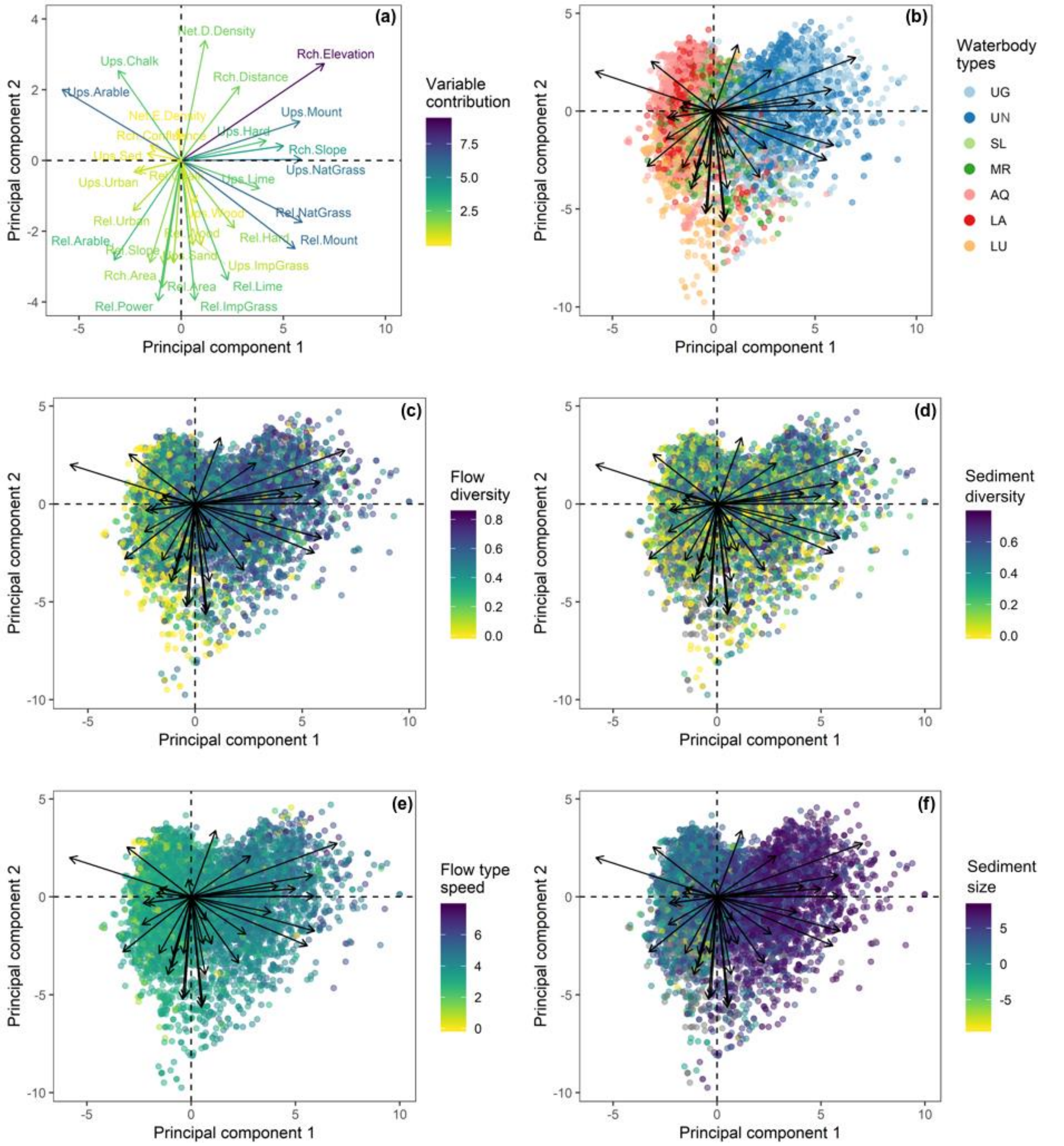


Figure 6.4. *PCA results on all control variables: (a) PCA loading plot of 34 variables, distribution of RHS sites on PCs 1 and 2 coloured by (b) waterbody type and (c-f) habitat indices.*

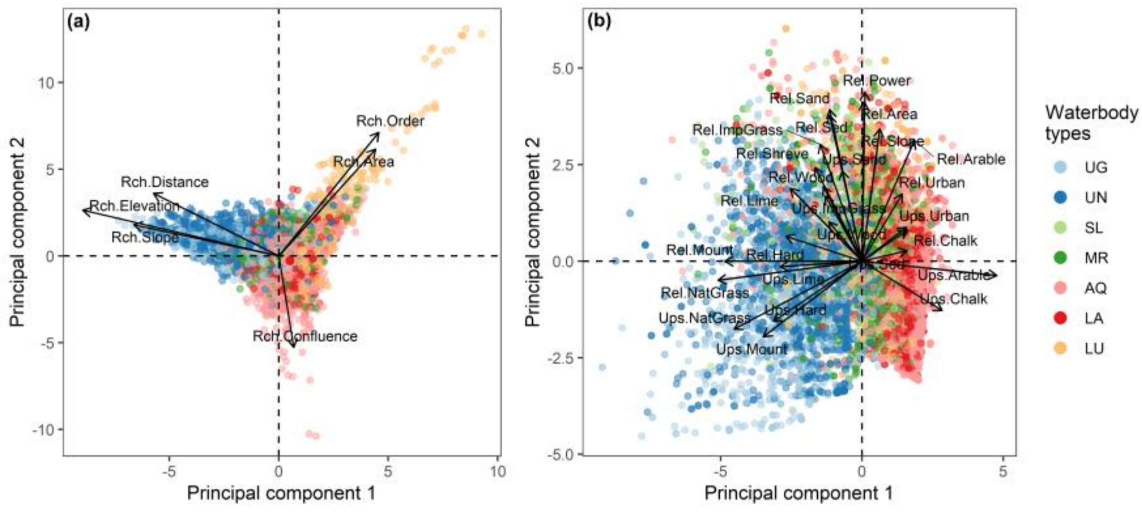


Figure 6.5. *Biplots of PCA results on (a) reach and (b) upstream and confluence control variables. Colours indicate waterbody types (see **Table 6.2** for abbreviations).*

Response to the PCs is not consistent between the habitat indices. For example, the reaches with the most homogeneous flow diversity tend to vary among the slower flow type speeds (**Figures 6.4c and e**) and sediment diversity shows no relation to sediment size or the PCs (**Figures 6.4d**). Therefore, it is critical to consider the habitat indices separately rather than as singular gradients as their response to catchment-level effects will vary, despite some common gradients.

6.3.2 Single regression tree

Average habitat type indices (flow type speed and sediment size) are predicted more accurately than the diversity indices (**Table 6.3**). The first few splits on each tree are determined by waterbody type, separating upland or lowland sites from the others (**Figure 6.6**). Most habitat indices, other than sediment diversity, also separate midland types (mid-range and seasonal) from other types in a secondary split. Sediment size retains the aquifer waterbody type in the midland types, likely due to the coarser grainsizes associated with permeable geology (Berrie, 1992).

Generally, habitat index values are highest in upland waterbodies declining towards lowland waterbody types (**Figure 6.6**, Section 3.2.3). Yet there is wide variation in habitat indices within each waterbody type even though values differ significantly between types (**Figure 3.3**). These trees show why this may be the case.

In upland waterbodies, average habitat indices are primarily driven by variations in reach slope, elevation and stream order with few splits needed to capture variation (**Figures 6.6c and d**). However, more variation in diversity indices is explained by confluence and upstream variables in upland waterbodies, than for average habitat type indices, reflecting variations in upstream inputs from the catchment (**Figures 6.6a and b**). For example, very low flow diversity and sediment diversity in upland waterbodies (i.e. 0.2 less diversity than other upland sites) is present when a site has >11% limestone geology upstream or a large difference in Shreve magnitude between tributaries respectively.

In lowland waterbodies, the lowest habitat type index values are associated with upstream geology (e.g. sediment size is coarse sand, **Table 2.3,** when there is a >83% total difference in sandstone geology between upstream tributaries) or anthropogenic land covers upstream (e.g. silt sediments, **Table 2.3c**, and dry to no perceptible flow, **Table 2.3a**) (**Figures 6.6c** and **d**). Similarly, flow diversity reduced by 0.18 compared to sites with similar catchment-level effects, when there was 1% urban land cover upstream (**Figures 6.6a**), which is likely due to the widespread homogenisation of channels in urban centres (Walsh *et al.*, 2005).

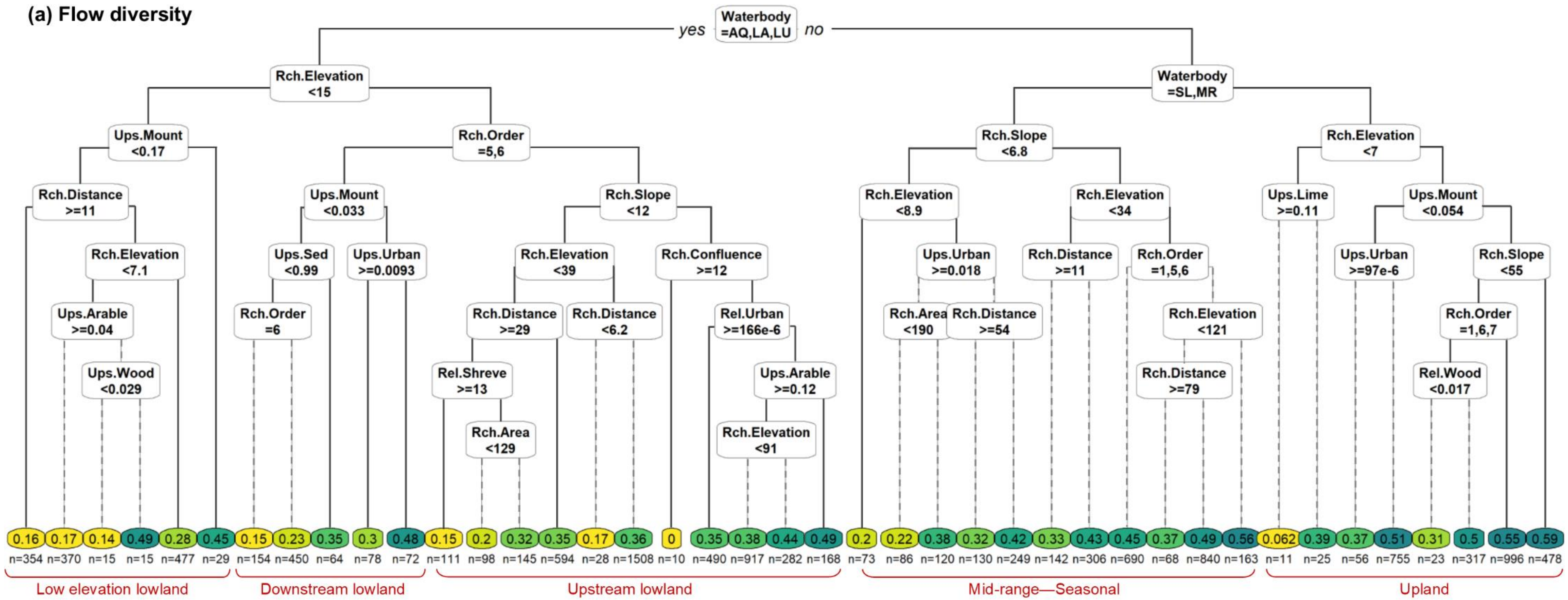


Figure 6.6. *Pruned regression trees for each habitat index (a-d). Left branches indicate that the condition at the split is true and right branches indicate the condition is false. Predicted habitat index values with corresponding colours are the leaves with the number of RHS sites in each leaf indicated below. Solid lines indicate branches which are retained by the more conservative method of pruning and so are the most critical.*

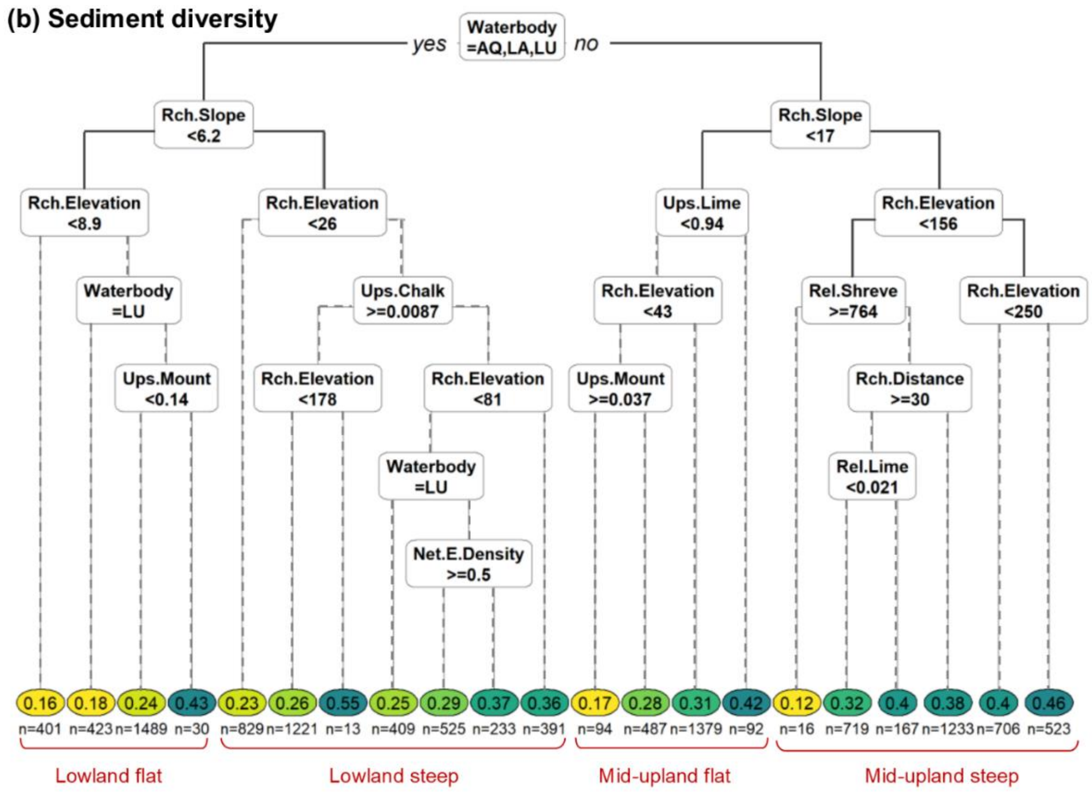


Figure 6.6. (continued). *Pruned regression trees for each habitat index.*

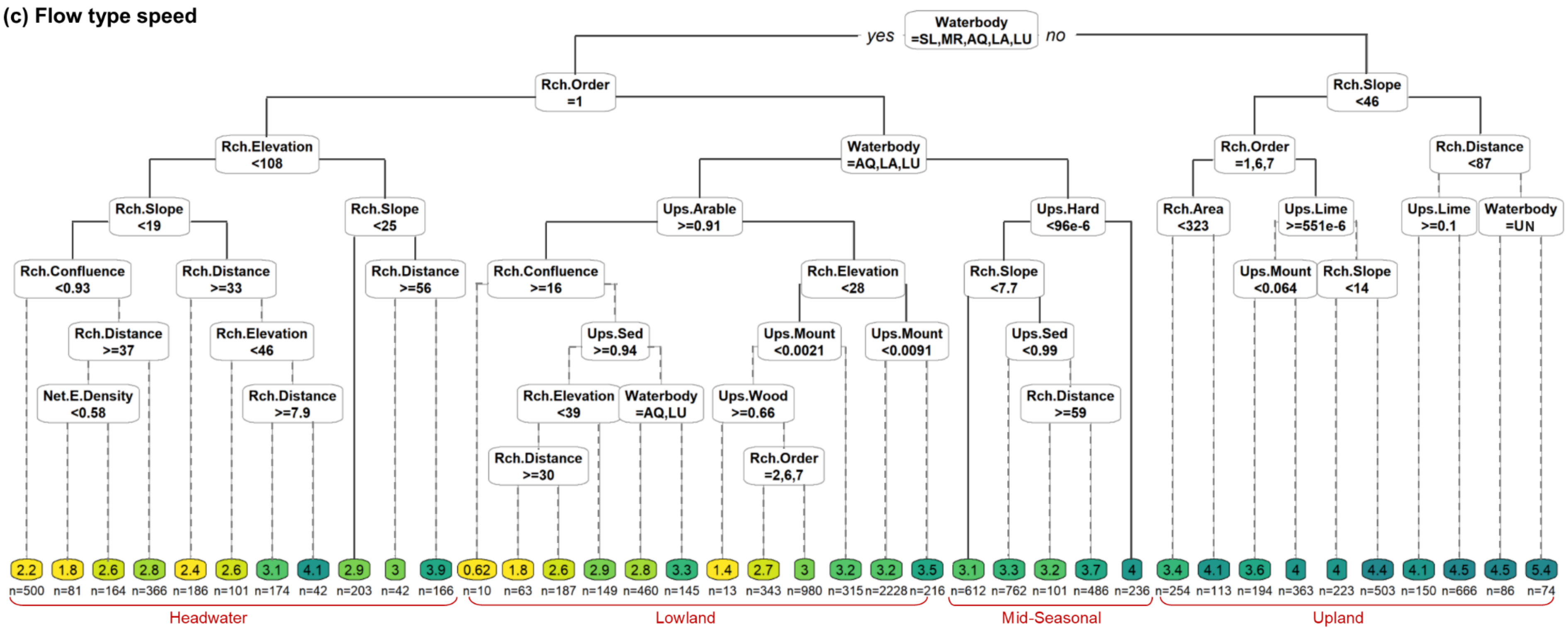


Figure 6.6. (continued). *Pruned regression trees for each habitat index.*

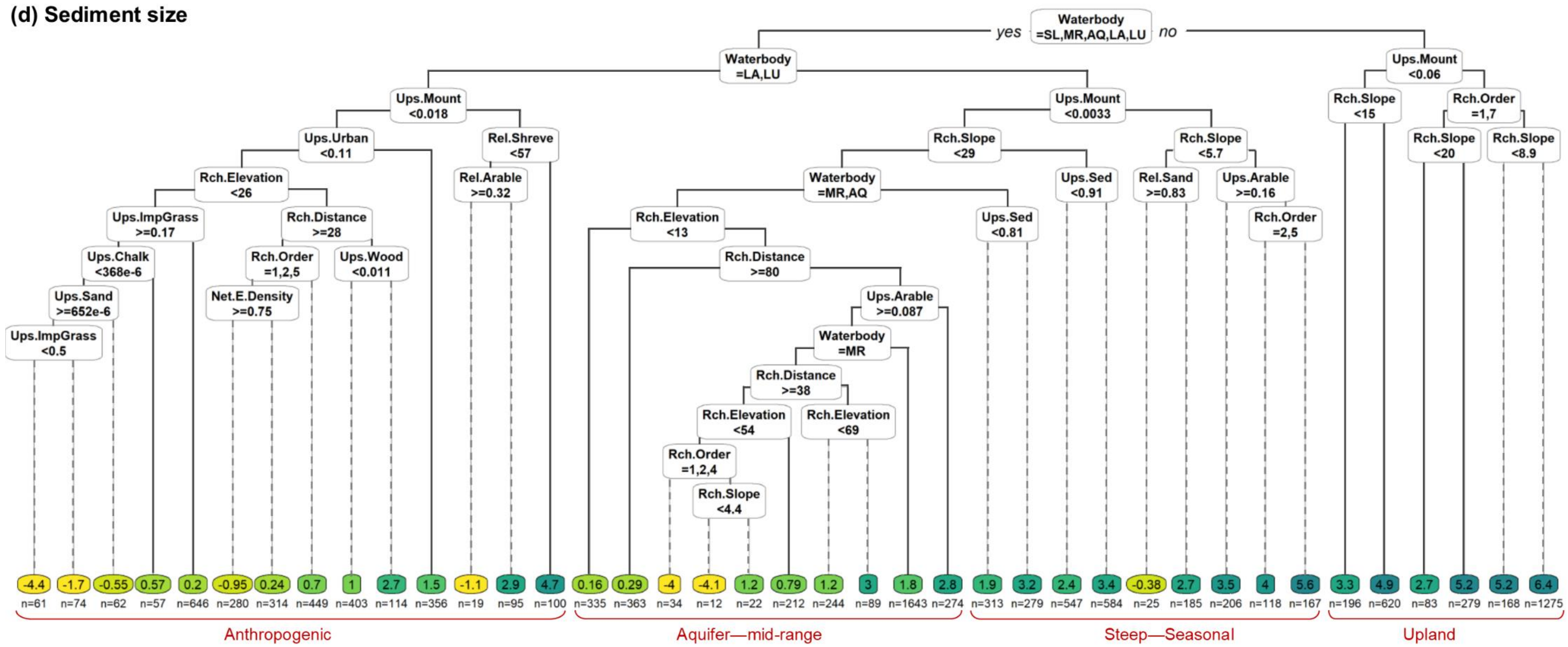


Figure 6.6. (continued). *Pruned regression trees for each habitat index.*

Alternatively, particularly high habitat type index values in lowland catchments occur at higher elevations, slopes or distances for flow type speed and with large differences in upstream tributary size for sediment size (**Figures 6.6c** and **d**). For habitat diversity indices, high values were associated with upland properties upstream, particularly mountainous and woodland land covers (**Figures 6.6a** and **b**). For example, sites with >3% woodland land cover upstream increased flow type diversity by 0.35, which may be due to the introduction of large wood to create more diverse biotopes (Cashman, 2014).

Network variables have little influence, aside from network elevation density in lowland waterbody types, with high density reducing sediment size and diversity. However, when there is a headwater stream in a network dense area, there is a 0.8 increase in flow speed index (**Figure 6.6c**). This may reflect increased coupling of catchment to channel where the landscape is more dissected by a dense river network.

The single regression trees illustrate the hierarchical nature of catchment-level effects, specifically that different variables are important within different waterbody types. They also show that while the key upland-lowland gradient is consistent between habitat indices, the addition of variables explaining upstream contributions helps to identify anomalous or extreme values.

6.3.3 Boosted regression tree models

6.3.3.1 Assessing model performance

BRT models have more accurate predictive performance than the single regression trees (**Table 6.3**). Predictive measures based on the CV training data are more conservative than those calculated from the final predicted values of the model (Waite *et al.*, 2019; **Table 6.3**) with a previous study stating that BRT R^2 values >0.3 indicated unsatisfactory model accuracy; 0.3–0.5, satisfactory accuracy; 0.5-0.6, good accuracy and; ≥0.6, very good accuracy (Rankinen *et al.*, 2019).

Table 6.3. *Summary and predictive accuracy of the tree models for each habitat index. Predictive accuracy calculated by testing final predicted values against observed values for (a) the single tree models and (b) BRT models. Training data values are tested against validation data values from the CV procedure for the BRT models to avoid inflated predictive accuracy (c).*

	Observed value range	BRT predicted value range	Number of trees	(a)Single tree final values R^2	(b) BRT final values R^2	(c) BRT CV values R^2
Flow diversity	0 - 0.84	-0.05 – 0.70	3400	0.20	0.43	0.28
Sediment diversity	0 - 0.82	0.01 – 0.58	2600	0.08	0.29	0.12
Flow type speed	0 - 7.9	0.13 – 6.62	5850	0.26	0.57	0.36
Sediment size	-9 - 8	-7.9 – 7.89	5250	0.41	0.65	0.47

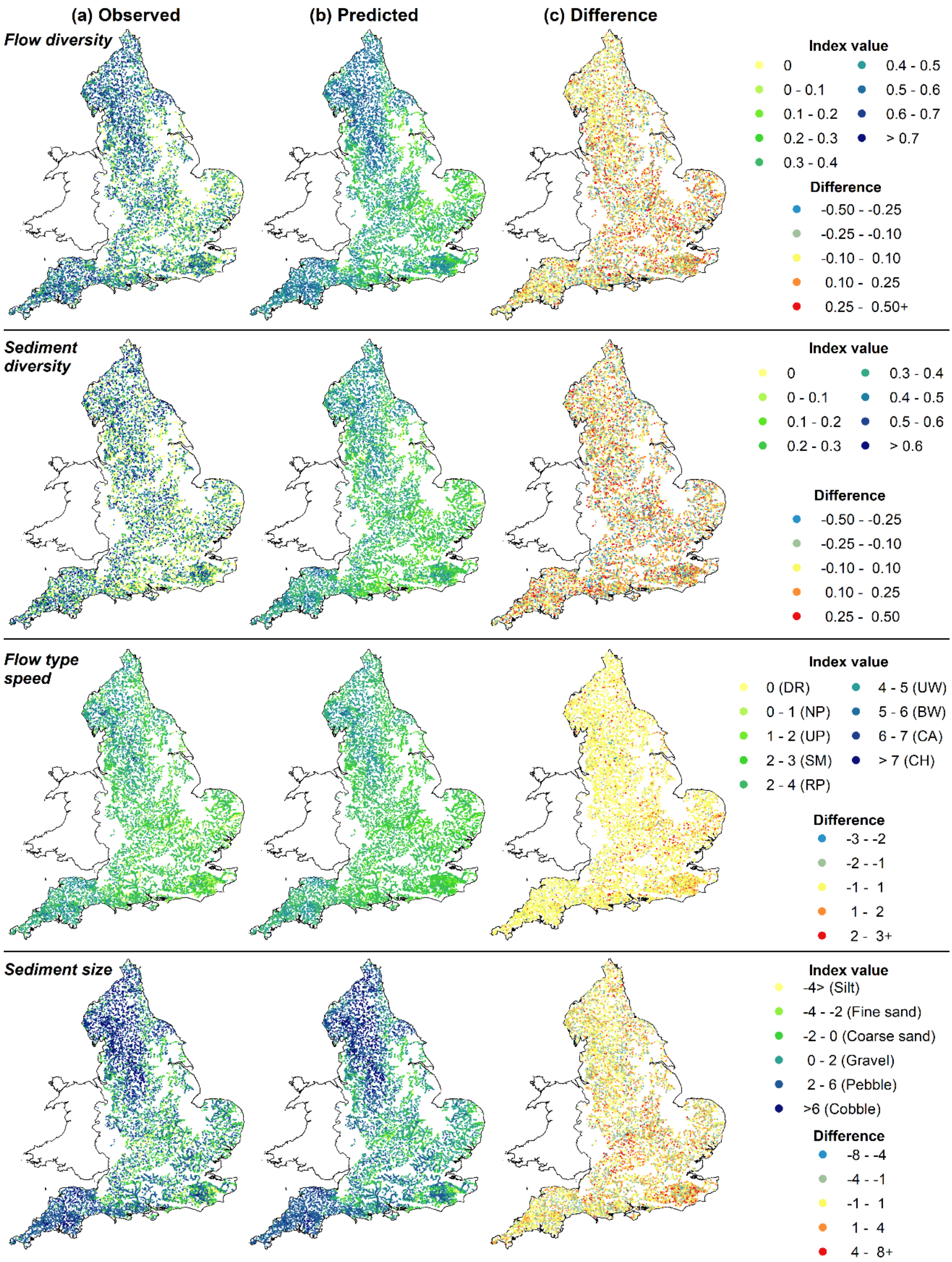


Figure 6.7. *Map of RHS sites used in BRT modelling indicating the (a) observed and (b) predicted habitat index values. (c) The difference between the observed and predicted values, reflecting sites where values are over or under predicted. See* ***Table 2.2*** *for flow type abbreviations.*

The diversity models have the lowest predictive capability (flow diversity CV $R^2 = 0.28$; sediment diversity CV $R^2 = 0.12$), over and under predicting with no clear pattern across the country (**Figure 6.7c**). The BRT models struggle to predict the extreme diversity values (**Table 6.3**), particularly homogeneous sites. This is because there is wide variation in the observed diversity indices across the country (**Figure 6.7a**). However, predicted diversity indices are tied to the dominant upland-lowland gradient (**Figure 6.7b**) which is exhibited in most of the control variables (**Figure 6.4**).

Average habitat type indices have more accurate predictions (flow type speed CV $R^2 = 0.36$; sediment size CV $R^2 = 0.47$) as the observed indices better reflect the upland-lowland gradient shown in the PCA (**Figure 6.4e** and **6.4f**). These models also struggle to predict the most extreme values (**Table 6.3**), particularly the dry flow types and silt sediment sizes in the aquifer waterbodies in central England and the seasonal waterbodies in the Weald in the south-east. They also under-estimate the extent of chute flow types and cobble sediments in the upland regions of England (**Figure 6.7**).

6.3.3.2 Variable importance and habitat response

The contribution of the control variables to the BRT models vary (**Figure 6.8**). Waterbody type and reach variables are the most important in all models, which is consistent with the findings from the PCAs (**Figure 6.4a** and **b**) and single regression trees (**Figure 6.6**).

Habitat index values decline as waterbody types become lowland. The diversity index models show a step change between upland-midland types (UG-MR) and lowland types (AQ-LU) (**Figure 6.9a and 6.9b**) whereas average habitat index models have a step change between upland types (UG-UN) and the other waterbody types (SL-LU) (**Figure 6.9c and 6.9d**). There are a few anomalies, with greater sediment diversity in lowland arable (LA) waterbodies, and coarser sediments in aquifer (AQ) waterbodies compared to other lowland types. These results reflect the primary splits in the single regression trees (**Figure 6.6**). Waterbody type may have high predictive power as it is a categorical variable.

Reach elevation is the second most important variable in all models (**Figure 6.8**), exhibiting a steep increase in habitat index values followed by a gradual increase or plateau in values between 50 and 200m (**Figure 6.9**). Other reach variables are frequently in the top ten important variables in all models (**Figure 6.8**), with fitted values of habitat indices increasing with slope and the distance from mouth or declining with upstream area and stream order for the diversity indices and with distance to confluence (**Figure 6.9**).

Network variables do not contribute meaningfully to the single regression trees but have a higher contribution to the models (2.8-4.9% importance) than many of the upstream and confluence variables (**Figure 6.8**). This suggests that network density improves model performance for poorly fitted observations in the model. There is a broad range in fitted values in the network partial dependency plots, but there are steep increases in fitted functions with a slight increase in network distance density (**Figure 6.9**).

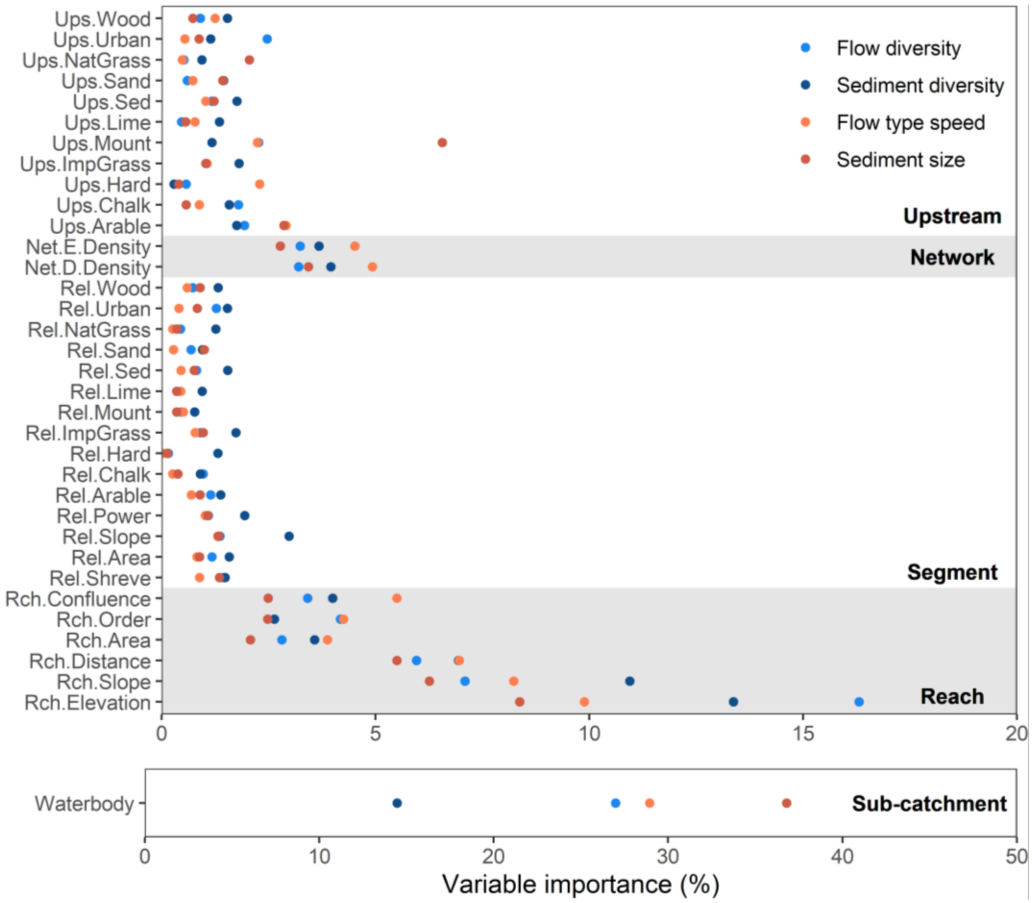


__Figure 6.8.__ Contribution of each control variable to the model for each habitat index. The x-axis on for waterbody type is on a different scale.

Surprisingly, upstream variables tend to not contribute substantially to the models compared to waterbody type (**Figure 6.8**). However, certain land cover variables are highly important (**Figure 6.8**): urban for flow diversity (contributing 2.5%), arable for flow type speed (contributing 2.9%) and mountain and arable for the sediment size model (contributing 6.6% and 2.9% respectively). Flow diversity exhibits a step decline with 60% urban cover in the upstream buffer (**Figure 6.9a**). Flow type speed and sediment size gradually decline with of the amount of arable land upstream (**Figure 6.9c** and **d**) and sediment size increases dramatically with the presence on mountainous land cover (**Figure 6.9d**). For the less important variables, effects are relatively consistent between habitat indices. For example, the presence of semi-natural grass or hard rock upstream induces a steep increase in habitat index values, the presence of chalk upstream causes a step decline in flow indices and increasing sedimentary or sandstone rock cover induces a slight decline in sediment indices (**Figure 6.9**).

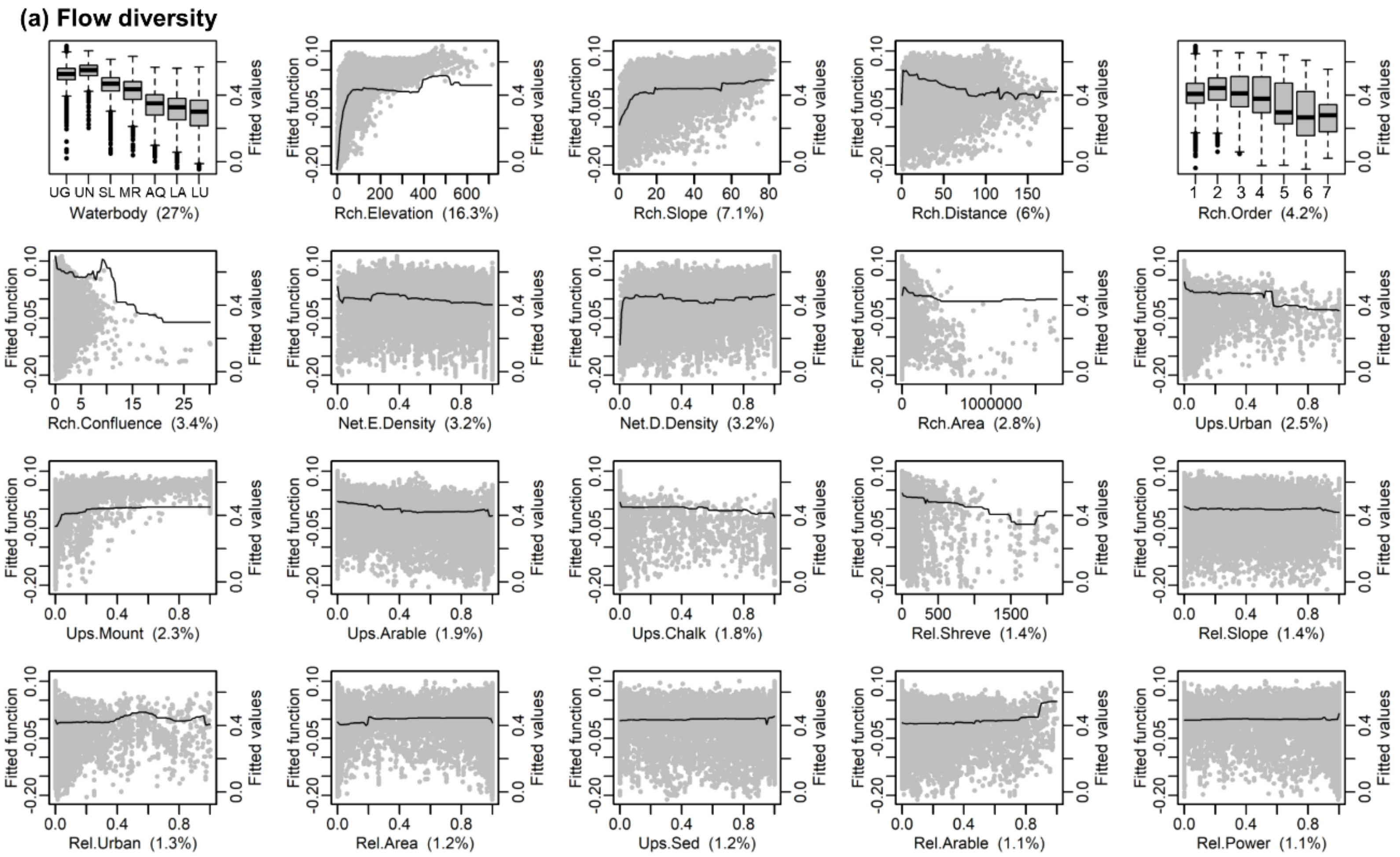


Figure 6.9. *Partial dependency plots for BRT models of habitat indices (a-d) with fitted habitat response function scaled on a normalised and centred (primary y-axis). Fitted habitat response values on habitat index scale (secondary y-axis). Contribution of each variable to the model in brackets. Abbreviations in* ***Table 6.2****.*

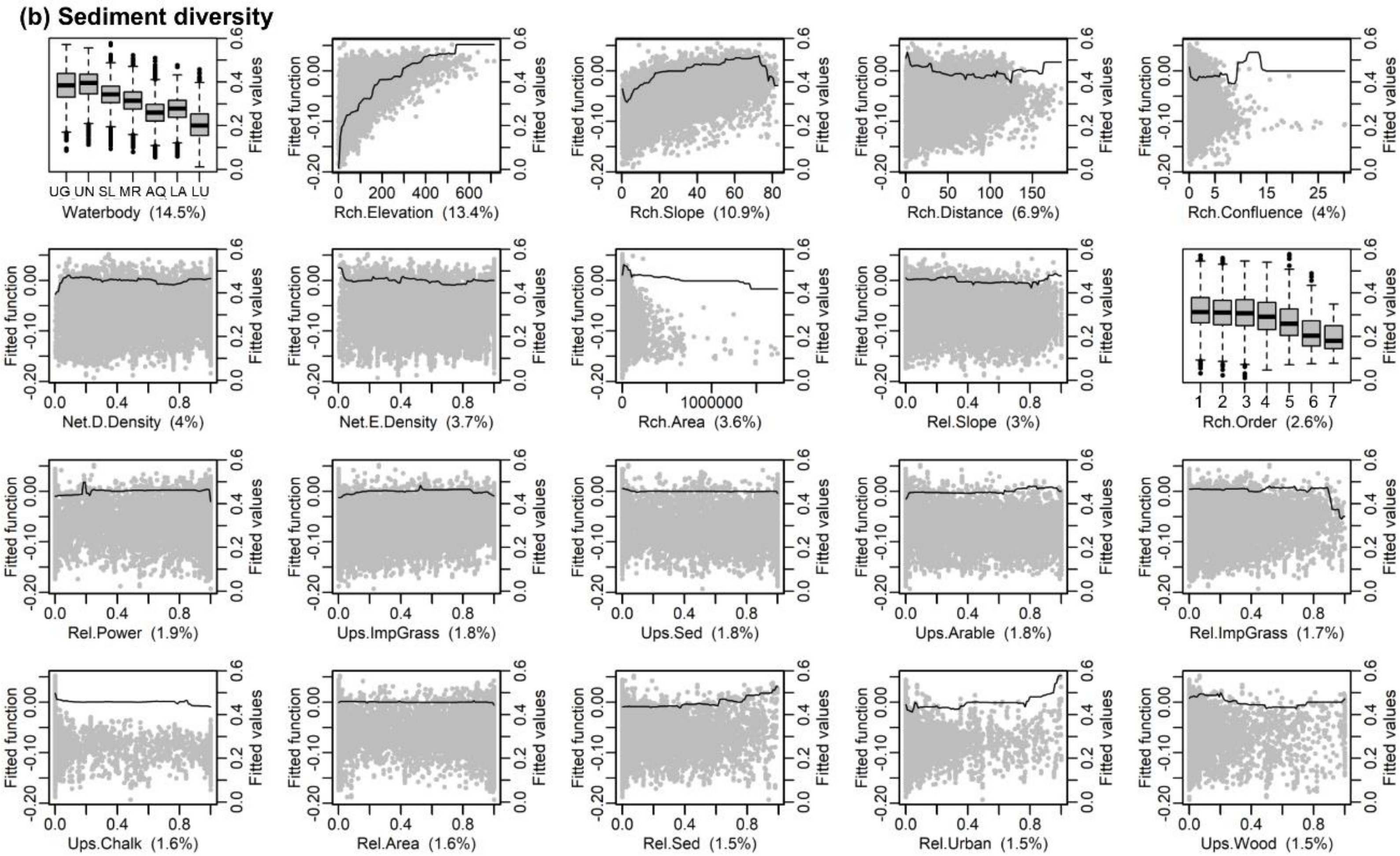


Figure 6.9. (continued). *Partial dependency plots for BRT models of habitat index (d) sediment diversity.*

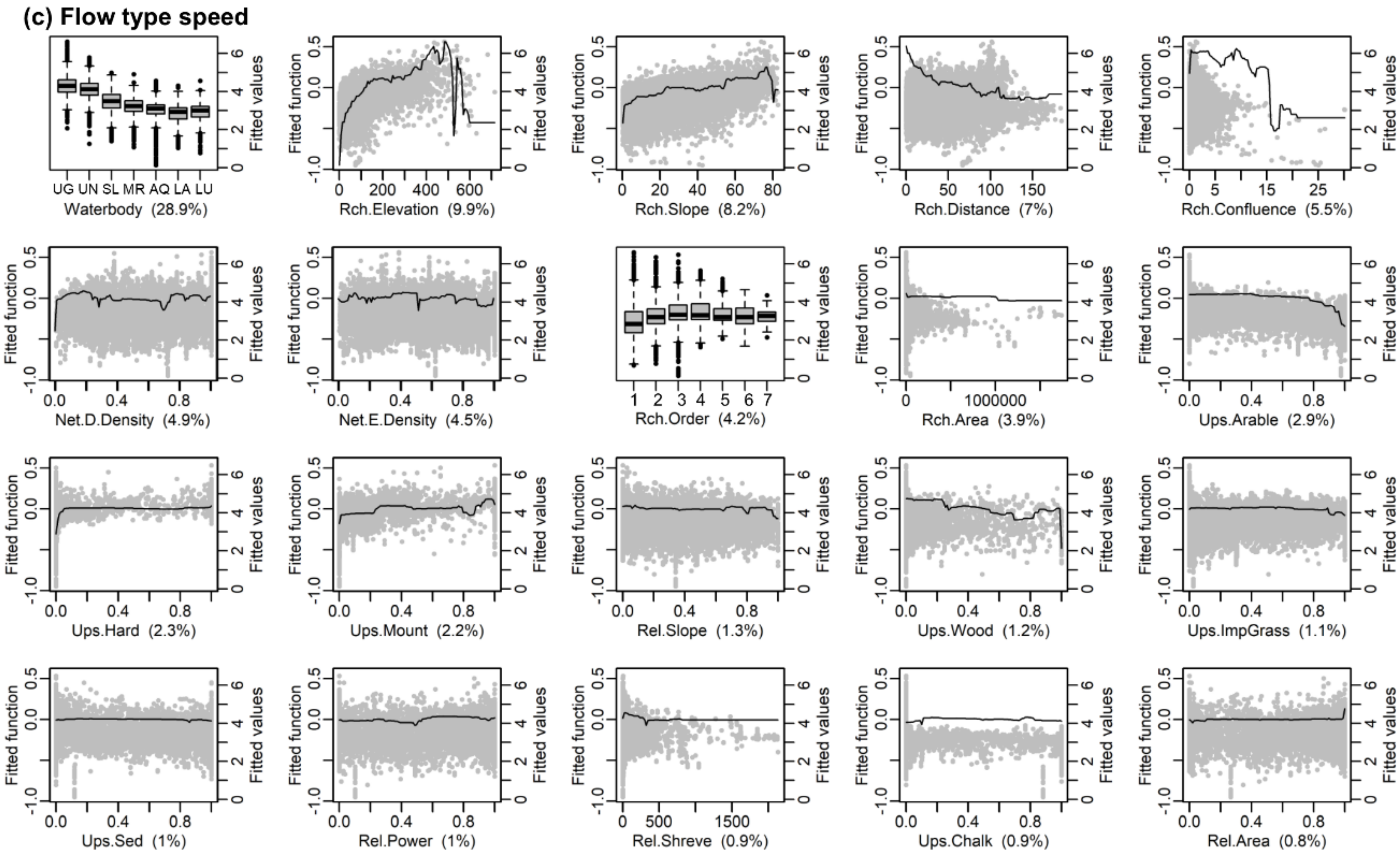


Figure 6.9. (continued). *Partial dependency plots for BRT models of habitat index (c) flow type speed.*

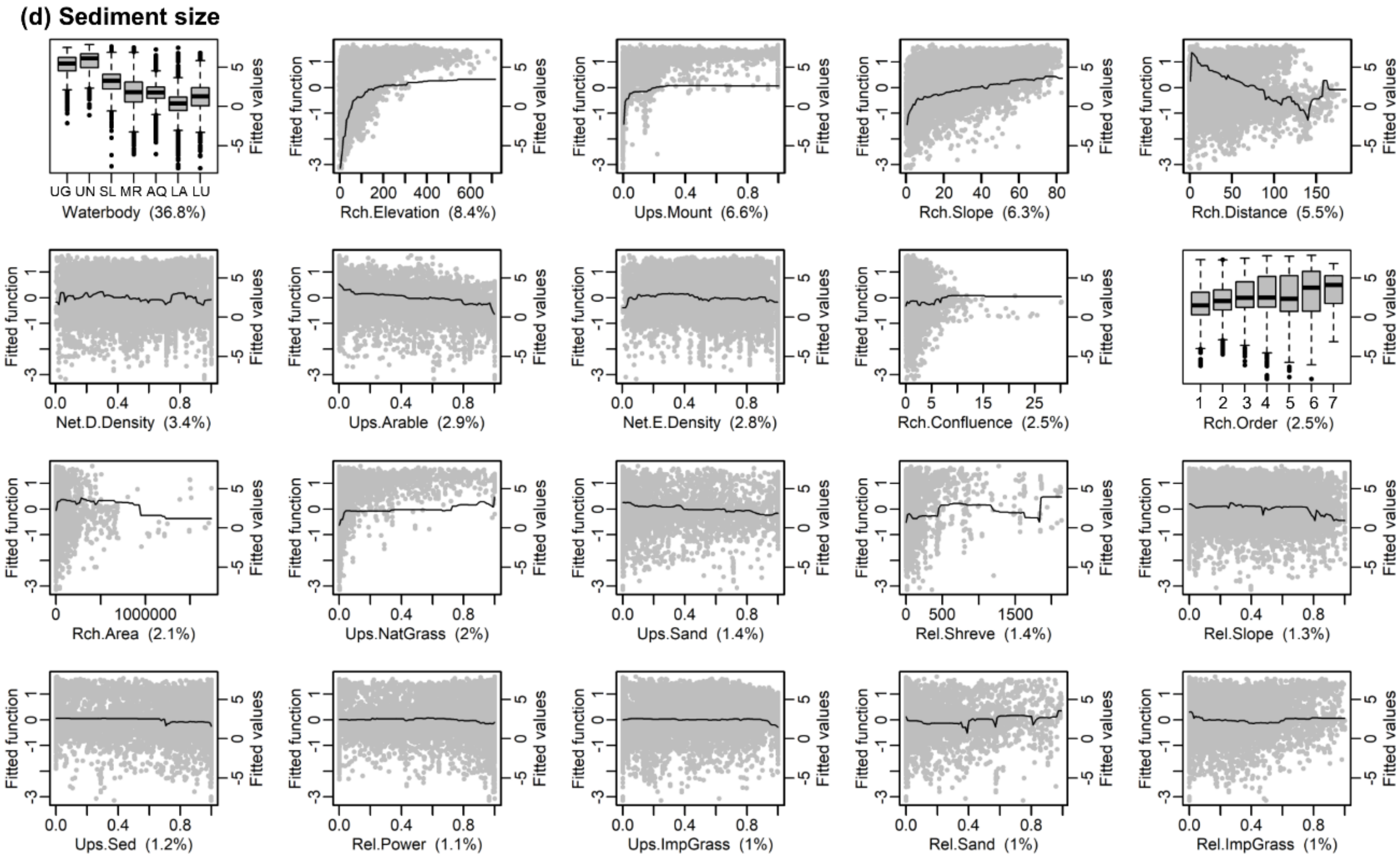


Figure 6.9. (continued). *Partial dependency plots for BRT models of habitat index (d) sediment size.*

Confluence variables are more important for diversity indices than average habitat type indices (especially sediment diversity) (**Figure 6.8**). This may be due to the wide spatial variation in these indices not captured by the upland-lowland gradient (**Figures 6.4c** and **6.4d**; **Figure 6.7a**) that may be partially explained by the different contributions from upstream tributaries. Hydromorphological confluence variables are often more important than upstream variables for all indices (**Figure 6.8**). Gradual declines in fitted values for habitat indices are present where there is high differences hydromorphological variables as tributaries converge, especially for relative slope (**Figure 6.9**). However, at the highest differences in slope and Shreve magnitude there is an increase in diversity values and sediment size.

For the land cover and geology confluence variables, there is wide variation in fitted values where there is no presence of the land cover upstream. The trends once individual land or geology percentages are present (i.e. >0%) are discussed here. Often the most important confluence variables are also important upstream variables (**Figure 6.8**). Flow diversity is influenced by the anthropogenic confluence variables with increases in diversity where there is a 20% and 80% total difference in urban and arable land covers between tributaries respectively. Sediment diversity shows a similar pattern with urban land cover and increases in diversity with high difference of sedimentary rock. Sediment size increases with large differences in sandstone and improved grassland between upstream tributaries (**Figure 6.9**).

6.3.3.3 Interrogation of interactions

Interactions between reach variables are present for all habitat index BRT models (matrices of interaction values in **Appendix 6E**), especially between reach elevation and distance from mouth. All 3D partial dependence plots show that high reach distance and low elevation equate to the lowest habitat index values (**Figure 6.10i**). This causes the baseline response of habitat indices to other variables to systematically increase with elevation and slope or decrease with distance. For example, the habitat response to the amount of semi-natural grassland upstream gradually increases with slope in the sediment diversity model (**Figure 6.10biii**).

Strong interactions with waterbody type are also frequent with variables affecting habitat indices differently in different waterbody types, with upland types often exhibiting a higher baseline habitat response than lowland types (**Figure 6.10iv**). For example, in upland waterbodies flow type speed increases more steeply with slope, and sediment size decreases more gradually with upstream arable land compared to lowland waterbody types (**Figure 6.10civ** and **6.10div**). For diversity indices, the response differs so greatly between waterbody types that habitat responses overlap. For example, flow diversity declines most steeply with relative Shreve magnitude in midland types, dropping below minimum flow diversity responses in the lowland waterbodies (**Figure 6.10aiv**). This is also the case for sediment diversity, where after 80% difference in slope between tributaries, large urban, lowland arable, mid-range and upland non-grass waterbodies have the steepest increases in diversity compared to other waterbody types (**Figure 6.10biv**).

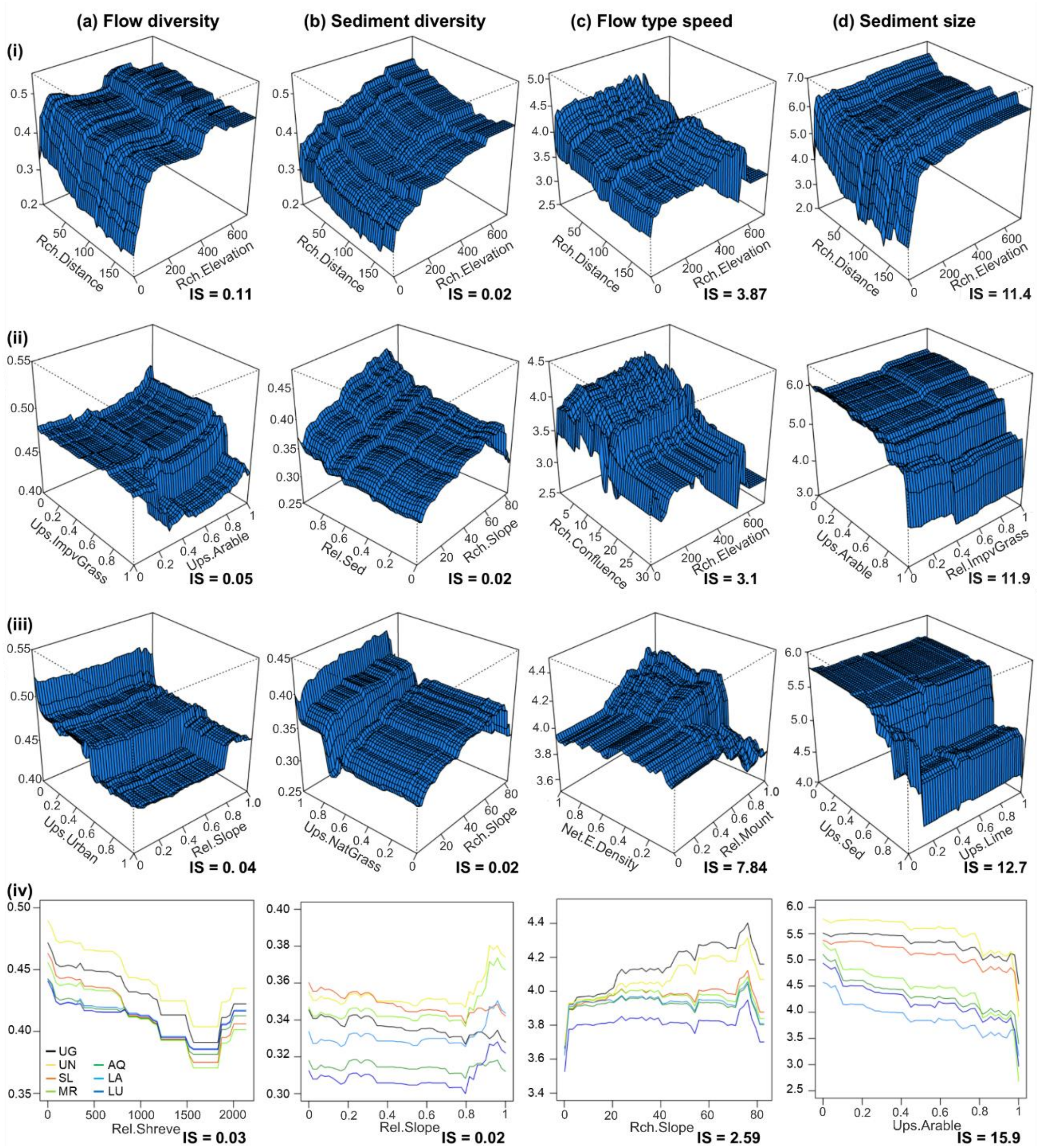


Figure 6.10. *Partial dependency plots of the strongest interactions between variables in the BRT models for each habitat index (a-d). y-axis of each pane is the fitted function of the model on the scale of each habitat index. Interaction strength (IS) indicated in bottom-right of each pane and is model dependent so cannot be compared between models. Interactions with categorical variable (i.e. waterbody type) are presented as two-dimensional graphs. Roman numerals are to aid citation of Figure in text.*

Network variables show the strongest interactions in average habitat indices models, exhibiting interactions with selected reach, confluence and upstream variables. The strongest interaction is in the flow type speed model. In this case, where the amount of mountainous land upstream is similar between the tributaries, network density has little effect on flow type speed. However, where there is >50% difference in mountainous cover, flow speeds increase at densities >0.5 but decline below this threshold (**Figure 6.10ciii**).

Upstream variables also often show interactions with confluence variables, particularly the anthropogenic variables. For example, where there are high percentages of arable land upstream, habitat indices decline, but there are dramatic step-change reductions in flow diversity and sediment size when there are also high percentages or high differences in improved grassland at a site respectively (**Figure 6.10aii and 6.10diii**). Similarly, where there is over 50% urban cover upstream and incoming tributaries have a similar slope, flow diversity is low. However, if the difference in slope is high, flow diversity is raised to near pre-urban levels (**Figure 6.10aiii**).

Geological variables also have interactions with other variables in the sediment models. For example, the positive effect of reach slope on sediment diversity is strongest where there is a high difference in sedimentary rock between tributaries (**Figure 6.10bii**). Similarly, sediment size declines with the amount of upstream sedimentary rock, but when the amount of sedimentary rock cover is over 80% and limestone is over 20%, there is a more dramatic decline in sediment size (**Figure 6.10diii**).

6.4 DISCUSSION AND SYNTHESIS

There are numerous mentions in this thesis of the spatial hierarchy within the river system, with small fluvial units nested cumulatively within larger units (**Figure 2.1**). This chapter has identified a nested structure of controls upon physical habitats within reaches. High-level controls on physical habitats reflect upland-lowland and upstream-downstream gradients in reaches, key properties of the river system (Section 6.4.1). However, within this high-level structure there is still wide variation in the types and diversity of physical habitats present within a reach. The variation is explored in numerous methods in this thesis, through the exploration of multiple catchment controls, network topology and tributary influences, all of which contribute to the catchment-level effect. This chapter highlights that low-level variation is driven by external upstream controls (Section 6.4.2) the influence of which varies along the high-level gradient.

6.4.1 High-level upland-lowland and upstream-downstream gradients

There is a clear upland-lowland gradient present in habitat indices evident from the PCA results (**Figure 6.4**) that has been identified in other studies using PCA to explain or predict distributions of channel features using subsets of the RHS dataset (e.g. Jeffers, 1998; Vaughan *et al.*, 2013). The primary principal components (PCs) in previous studies were driven by reach variables of altitude, slope (Jeffers, 1998) and stream power (Vaughan *et al.*, 2013). Here, the upland-lowland gradient is driven by many variables at multiple spatial levels (**Figure 6.4a and 6.4b**) and the reach-level in isolation does not reflect the distribution of sites as well as the multi-level characteristics (**Figure 6.5a**). This highlights the utility of using multiple variables to explain RHS feature distribution as opposed to a limited number of variables (e.g. Vaughan *et al.,* 2013), that may be good predictors of habitat features but do not explain the key controls on the system (Mac Nally, 2000).

The waterbody types also reflect the upland-lowland gradient (Section 3.2.4.3; **Figure 6.4b**), representing the primary split in the single regression trees (**Figure 6.6**) and the most important variable in the BRT models (**Figure 6.8**). This may be because waterbody type captures not only the upland-lowland gradient but regional controls on reaches that have before been shown to influence RHS habitat features (Harvey *et al.*, 2008b; Vaughan *et al.*, 2013; Naura *et al.*, 2016; **Table 6.1**). These additional variables reflect catchments with differing geologies, topographies, climate and land cover whilst at similar positions along the upland-lowland gradient.

While the high-level upland-lowland gradient captures regional trends in boundary conditions (Schumm and Lichty, 1965; **Figure 2.1**), there is variation within similar waterbody types (Section 3.2.4.3). This is because in the waterbody typology, a reach at the outlet of a waterbody is classified the same as the headwater reach, yet there are an array of processes occurring from steep sided valleys upstream to wide floodplain landscapes that create a range of habitats within a single waterbody (Schumm, 1977; Vannote *et al.*, 1980; Church, 2002; **Figure 2.2**). The second high-level gradient (**Figure 6.4a**) captures variation *within* catchments rather than *between* catchments reflecting the upstream-downstream dimension found in other studies (e.g. Jeffers, 1998; Vaughan *et al.*, 2013).

Figure 6.4b shows that the waterbody types do not respond to this secondary gradient except for large urban waterbodies where large upstream catchment area was a key descriptor of this type (Section 3.2.3.2). Reach-level characteristics, on the other hand, reflect the upstream-downstream gradient. They are the second most important level of variables for explaining physical habitats (**Figure 6.8**) because in combination with waterbody type, they capture both regional and internal variation in catchments. This is evidenced by the single regression trees that show that the first splits in the tree are between waterbody types and subsequent splits separate upstream from the downstream sites within types based on reach-level characteristics (**Figure 6.6**).

The amount of variation explained by reach variables varies along the upland-lowland gradient. In upland waterbodies, variation in habitat index values may be explained by reach elevation and slope, with downstream sites exhibiting lower habitat values than other upland sites. However, variation in the midland and lowland waterbodies may not be captured so simply. In these cases, other factors from the upstream network offer a more accurate reflection of within catchment variation than upland-lowland gradients within catchments.

6.4.2 Low-level influences from the upstream network

6.4.2.1 Upstream thinking

Within catchment variation in habitat type and diversity is not only influenced by high-level upland-lowland and upstream-downstream gradients, but also the properties of the upstream network. For example, low habitat index values are associated with increasing amounts of upstream arable (**Figure 6.9a**, **6.9c** and **6.9d**) and urban land (**Figure 6.9a**), reducing habitat values in comparison to similar reaches across a range of waterbody types (**Figure 6.6**). The negative effect on habitats is because anthropogenic land covers are linked to increases in fine sediments (Wharton *et al.*, 2017), over-widening, straightening and dredging practices that create homogeneous channel environments (Sear *et al.*, 2003; Walsh *et al.*, 2005). However, the percentage of anthropogenically modified land in the upstream network required to cause negative effect differs between urban and arable land covers. For example, only a small percentage of urban land upstream causes habitat indices to decline (Figure 6.6) however, habitat indices gradually decline with increasing amounts of arable land upstream (**Figure 6.9**). This observation concurs with a study on ecological quality in UK waterbodies, which declined sharply with 0-5% urban land cover upstream yet declined at a much slower rate with increasing arable land (Smith, 2015). Despite urban area taking up a smaller percentage of the UK than agricultural land (which occupies over 75% of UK land use; Khan *et al.*, 2013), urban land upstream exhibits an strong influence on both proximate and distant rivers (Paul and Meyer, 2001).

Upland catchment characteristics, especially mountain/heath/bog land cover, are also associated with sharp increases in habitat indices with only a small percentage present upstream (**Figure 6.8** and **6.9**). This positive effect of natural and upland characteristics on habitat features has been shown in other studies (e.g. Feld, 2004; Manfrin *et al.*, 2016) and this study highlights that the positive effect is especially high in lowland reaches that typically have low habitat index values (**Figure 6.6**). This is because upland areas in the UK have a higher coarse sediment supply and also greater sediment transport capacity (Raven *et al.*, 2010) to transport coarse sediments to reaches downstream, likely improving habitat conditions compared to similar sites without an upland influence. However, in upland waterbodies, specific upland characteristics have little effect as all sites in these waterbodies are influences by upland characteristics (**Figure 6.6**).

Despite these noted effects of upstream characteristics, they have little influence on the final BRT models compared to waterbody type (**Figure 6.8**). This may be because characteristics nearer to the stream potentially have a greater influence on the reach than characteristics in the distal regions of the network, however, a review shows that studies find contrasting results (Allan, 2004). For example, studies in the same region in Michigan showed land cover locally upstream of the reach had the greatest influence when sites were in catchments with similar land cover characteristics (Lammert and Allan, 1999) yet, when sites were in contrasting catchments, catchment-level land cover is more influential (Roth *et al.*, 1996). This suggests that both levels are important; the catchment-level for determining wider boundary conditions of the catchment, and local upstream characteristics for identifying site specific influences. This supports the findings of this chapter, identifying high-level and low-level controls on physical habitats. Much research on the influence of spatial level on reach features is focused on land cover (Allan, 2004) but here upstream geology is also influential in the single regression trees and in the BRT models (**Figures 6.6** and **6.9**). Geology has also previously been related to RHS reach features, particularly sediment size and bedform diversity (Emery *et al.*, 2004; Harvey *et al.*, 2008b; Naura *et al.*, 2016). These results highlight the importance of upstream thinking in studies on river reach functioning as distal influences upstream can be propagated downstream through the network.

6.4.2.2 Network topology impact dependent on catchment conditions

The network density metrics developed in Chapter 4 (Heasley *et al.*, 2019) are more important than many upstream and confluence variables in the BRT models (**Figure 6.8**) suggesting that the network density metrics may help describe variation not explained by the high-level gradients in the habitat indices. However, the results from Chapter 5 suggest that network density metrics may only have a significant influence under certain conditions and the interactions from the BRT model indicate under which conditions this may occur. For example, elevation network density only influences flow type speed if there is a difference in the mountainous land cover between upstream tributaries. Under these conditions, flow types are faster at high network densities and slower at low densities (**Figure 6.10ciii**). This is supports evidence from Chapter 5 and other empirical studies that shows confluences in the network have an influential effect on river reaches in high-energy environments (Benda *et al.*, 2004a; Rice *et al.*, 2001), whereas confluences in lower energy, modified landscapes have less of an impact (Singer, 2008).

The inclusion of network density improves the explanatory power of the model, but it is difficult to identify exactly how network density is influential. From the interactions in the model, it is suggested that network density variables have different effects on habitat indices if there are large differences in upstream properties, building on the work in Chapter 5. Further research to explore this in depth would be beneficial by exploring network density effects over a wider range of climates and conditions.

6.4.2.3 Importance of tributary heterogeneity

As shown above, the relative contribution from two tributaries as they join at a confluence can influence the effect of network density on habitat indices, but also other variables. Relative tributary characteristics, or confluence variables, are not very important in the BRT models (**Figure 6.8**) yet they show some of the strongest interactions with other variables (**Figure 6.10**). This demonstrates the importance of considering relative tributary characteristics as they can impact habitat substantially under certain conditions. For example, where there is a strong urban influence upstream, if an incoming tributary has a similar slope to the main channel, flow diversity is unaffected. However, a high slope tributary increases flow diversity to near low-urban levels (**Figure 6.10aiii**). High differences in slope between tributaries also increases sediment diversity in anthropogenically impacted waterbody types more than other types (**Figure 6.10biv**). This is because of the enhanced transport capacity of the steep incoming tributary creating a range of bedforms downstream of the confluence (Debnath *et al.*, 2019; Rice, 1998) that may otherwise be missing in the reaches with anthropogenic controls.

Dissimilarity in geology and land cover between tributaries also induces a positive habitat response. Where there is high dissimilarity in land cover and geology types that are shown to be detrimental to habitat (**Figure 6.9**), there is a positive effect on habitat indices. This suggests that the negative effect of the tributary with the detrimental characteristic is mitigated by the contrasting characteristics of the other tributary. This supports observations of ecological improvements where a natural stream joins a modified system (Sabo *et al.*, 2018). Certain characteristics can enhance this mitigating effect, for example, in steeper reaches dissimilarity in sedimentary geology has more of a positive effect on sediment diversity than in reaches with a lower gradient (**Figure 6.10bii**). There is little evidence for reduced habitat indices as a result of land cover or geology similarity. However, under certain conditions (such as where there is low network density, **Figure 6.10ciii**) dissimilarity in land covers that primarily have positive effects on habitats (e.g. mountain/heath/bog land covers) cause a decline in habitat indices. This supports evidence of a negative impact on ecology where poor quality streams join higher quality streams (Cooksley *et al.*, 2012).

Equally sized tributaries are often shown to have positive effects on habitat variables downstream (Benda *et al.*, 2004a; Best, 1987; Jones and Schmidt, 2018). There is evidence for this in the sediment size BRT model (**Figure 6.9d**) but some other indices show little evidence of this effect (**Figure 6.9**). The results of the BRT models support the findings from Chapter 5, that identified that dissimilarity between tributaries could cause positive or negative effects on habitats. The positive or negative effects were found to be influenced by relative tributary characteristics, although not strongly. This analysis shows that positive or negative effect of a tributary is not solely dependent on relative tributary properties, but also the characteristics of the reach and its upstream influences. This suggests that the simple probability matrix presented by Jones and Schmidt (2016), that predicts an abrupt change downstream of confluences if the tributaries are of a similar size and/or the landscapes are dissimilar, is too simplistic.

6.4.3 Challenges in predicting physical habitats with catchment-level effects

While the information derived from the tree models provides interesting results from an explanatory perspective, which was the objective of this chapter, the predictive power of the models is relatively low.

Both regression tree and BRT models for all habitat indices struggled to predict extreme values (**Table 6.3**; **Figure 6.7c**) partially because the regression is designed to reduce error rather than predict extremes but also because of natural variation and local influences. For example, dry flow type speeds may be a result of abstraction, a practice that is more prevalent in permeable geologies associated with aquifer waterbodies (Petts *et al.*, 1999) which is reflected in the predictions (**Figure 6.7b**). However, dry flows may also be a result of impoundments such as reservoirs which are not directly included in these models.

The habitat diversity models have particularly low predictive capacity with CV R^2 values of 0.28 and 0.12 compared to habitat type models CV R^2 values of 0.36 and 0.47 for flow and sediment indices respectively (**Table 6.3**). This may be because habitat diversity is more dynamic and responds to rapid discharge changes and channel vegetation (Padmore, 1997) compared to average habitat type that adjusts over time to dominant reach conditions during formative flow events (Bunn and Arthington, 2002). Also, both upstream and downstream sites may produce relatively homogeneous channel morphologies and habitats if dominated by solely erosional or depositional processes (Schumm, 1977; **Figure 2.2**), making diversity hard to predict from the dominant gradients in the variables (**Figure 6.4**). Therefore, as habitat diversity indices do not respond to the high-level gradients as clearly, unlike average habitat type indices (**Figure 6.4**), there are inaccuracies in prediction (**Figure 6.7**). Upstream and confluence variables capturing some of the additional variation not accounted for by the high-level gradients (**Figure 6.8**), however, homogeneous sites are particularly hard to predict (**Figure 6.7**), partly because homogeneous sites are not only located in relation to anthropogenically modified land covers, but also local human modifications not accounted for in this analysis.

While there are challenges, the objective of this chapter was to build an explanatory model to explain patterns of physical habitats using solely GIS-derived variables. This has been achieved by improving understanding of high-level influences on physical habitats although low-level variation could not be fully accounted for. Solely GIS-derived variables were used in the model so that the relationships identified could potentially be extrapolated to predict habitats at sites with no survey data. This would be beneficial to reduce the need for conducting expensive, time-consuming surveys nationally. However, even though the BRT models described in this chapter can explain high-level patterns in habitats, they may not be effective as predictive tools. Predictions at non-surveyed sites would likely have high inaccuracy as the models would predict broad patterns in physical habitats but lack detail at a local level. Although, the inclusion of variables at the local level from the RHS surveys (e.g. types of channel modifications) could improve predictive performance by capturing some local variation, the resultant model could not be used to predict physical habitats at un-surveyed sites. Therefore, an alternative application of the models is explored below to better understand how high-level gradients can be combined with local knowledge to help strategically target restoration practises.

6.4.4 River management application of multi-level tree model

This thesis has focused on explaining broad-scale patterns of physical habitats at a national level using GIS-derived data that represent catchment-level effects using data-science techniques. As stated previously, such broad methods are often focused on accurate prediction, however the focus of this chapter was on explanation. Therefore, here the broad-scale patterns identified are combined with local knowledge to explore how an individual site is similar to, or different from, other sites with similar catchment-level effects. Thus, the restoration potential of the site can be assessed by identifying how much intervention is necessary to make habitat at a site comparable with sites with similar influences. Similar applications of predictive regression models are presented by Vaughan *et al.* (2013), that compare sites with similar geomorphic conditions. However, the example presented here demonstrates the utility of explanatory models for performing similar tasks that include not only measures of reach-level geomorphic processes, but the broader catchment context. The example presented below is only intended to be illustrative.

To explore potential applications of the tree models, sites with similar catchment-level effects on flow type speed were identified if they were classified in the same leaf of the single regression tree (**Figure 6.11a**). The single regression trees are used rather than the BRT model, as the BRT model was designed as a predictive tool so individual trees are not intended to be extracted for this purpose (Brandon *et al.*, 2019). However, the first tree modelled by BRT (before subsequent trees were built to explain poorly modelled observations; Elith *et al.*, 2008) was comparable to the single regression tree that utilised the conservative pruning method to avoid overfitting (**Appendix 6A**). The tree groups sites based on high-level catchment controls, primarily waterbody type, and reach-level controls. Here, the flow type speed index is selected to identify sites with similar physical biotopes to the sites in question (**Figure 2.4**) but this methodology could be applied to any habitat index of interest. Three RHS sites were selected at random, each within a different leaf of the tree (**Figure 6.11a**), so the selected sites could be compared to other sites at similar positions along the upland-lowland and upstream-downstream gradients. The selected sites were on the rivers Wansbeck, Ise and Rase.

Site 1: Wansbeck

The Wansbeck site is a 4^{th} order stream draining a limestone watershed. Sites with similar catchment-level effects to the Wansbeck are upland sites with steep slopes (**Figure 6.11a**) that are distributed in the upland regions of England (**Figure 6.11b**). The Wansbeck is compared to these similar sites to identify how flow type speeds compare to other sites and observations from the RHS dataset are used to explain why this may be the case.

The average flow type speed of the Wansbeck site is lower than the average speed of similar sites (**Table 6.4**). According to the biotope matrix in **Figure 2.4** this indicates that the Wansbeck site conditions have an average 'run' physical habitat, whereas the average flow for other similar sites would indicate a faster, shallower 'riffle' physical habitat (Padmore, 1997; Rowntree, 1996). This is important as riffle habitats are critical for fish spawning and are often degraded (Plug *et al.*, 2013).

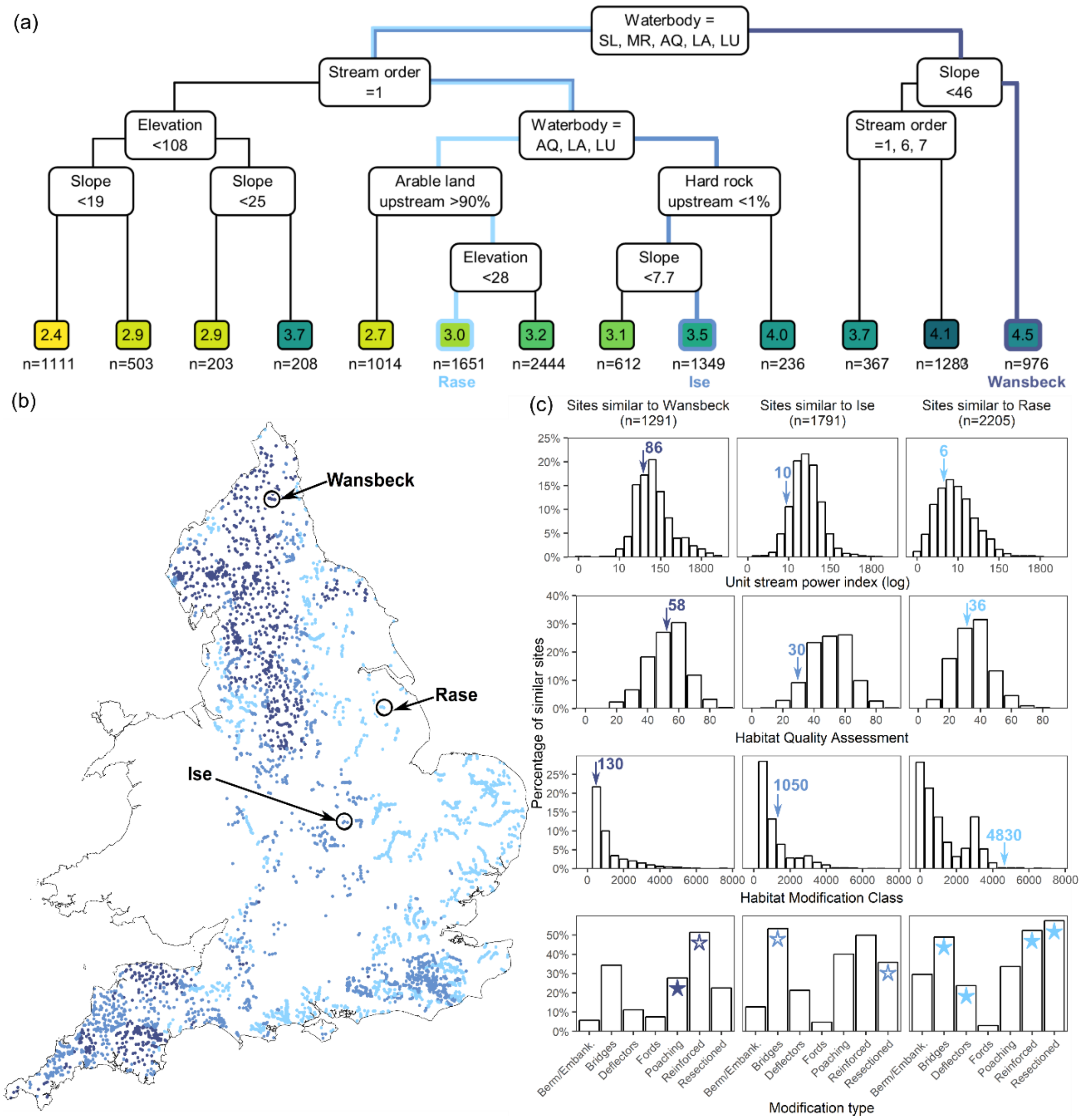


Figure 6.11*. Example of workflow to compare three RHS sites, the Wansbeck, Ise and Rase to sites with similar catchment-level effects on flow type speed. (a) Single regression tree of flow type speed from Figure 6.6c using the conservative pruning method, colours indicate the characteristics of each of the three sites. (b) Map showing the three RHS sites and sites with similar catchment-level effects. (c) Distribution of unit stream power, Habitat Quality Assessment, Habitat Modification Score and modification type values within the similar sites. The value for each of the three sites is indicated be the arrow. Stars indicate presence of a modification type at the site, if the star is filled in, the modification is extensive at the site.*

Aside from the lower flow type speeds, the Wansbeck itself has characteristics relatively typical of similar sites, with slightly lower stream power and HQA than other sites, and low modification score. However, the site has evidence of extensive bank poaching indicating livestock activity at the site and banks have been reinforced (**Figure 6.11c**). This indicates a loss of channel stability due to livestock activity that may have caused channel widening (as livestock trample the banks), incision (as livestock trample the cobble bed armouring, reducing resistance to vertical erosion) and increases in flow depth during peak flows (as livestock trample the soil increasing runoff) (Belsky *et al.*, 1999; Trimble and

Mendel, 1995). This may reduce the dominance of faster flow types, such as unbroken waves, at the site as the hydraulic control on surface topography is drowned out by deeper flows and the bed is broken up by livestock (Padmore, 1997). Limiting livestock access to the stream would be a relatively low-cost measure to reduce the effects of poaching, however, a field survey to determine if this is the only cause of the comparatively slower flow type speeds in the Wansbeck would be necessary.

Table 6.4. *Predicted flow type speed of all similar sites compared to the actual flow type and other habitat variables of the selected site. Average flow type speed and sediment size classes indicated in italics.*

Site	Predicted flow type speed of all similar sites	Actual habitat index values of the selected site				Bankfull width at site (m)
		Flow type	*Sediment size*	*Flow diversity*	*Sediment diversity*	
Wansbeck	4.5 rippled-unbroken wave	3.8 smooth – rippled	7.2 large cobble	0.62	0.32	5
Ise	3.5 smooth – rippled	1.6 no perceptible flow – upwelling	3.5 pebble	0.42	0.00	1.9
Rase	3.0 smooth	3.1 smooth	-1 sand	0.18	0.18	8.5

Site 2: River Ise

The site on the River Ise is a geomorphically homogeneous reach at a moderate elevation, draining a sedimentary rock watershed. Sites similar to the Ise are midland sites with little to no hard rock geology upstream and moderate slopes (**Figure 6.11a**). They are primarily located in the high elevation regions of the south west and south east, with some sites in central England (**Figure 6.11b**). Other similar sites are dominated by a faster, run habitat than the Ise (**Table 6.4**) which also has lower stream power and habitat quality, and higher modification score than similar sites (**Figure 6.11c**). In particular there is a bridge at the site and evidence of resectioning which likely resulted in channel deepening (Environment Agency, 2003) producing a predominantly deep, slow, pool habitat (Harvey *et al.*, 2008a). This may indicate that the Ise site has potential for restoration to increase faster, shallower flow types in the reach to promote habitat diversity.

Site 3: River Rase

The site on the River Rase drains a >6000 km^2 catchment area and is at a relatively low elevation. Sites similar to the Rase are lowland sites with less than 90% arable land upstream and low elevation (**Figure 6.11a**), located in the chalk aquifer and lowland areas of England (**Figure 6.11b**). The Rase itself has a relatively similar unit stream power and HQA to other sites but a much higher modification score with extensive evidence of bridges, deflectors, reinforcement and resectioning (**Figure 6.11c**). Despite these extensive modifications the Rase has a similar flow type speed compared to other sites (**Table 6.4**), potentially as a result of the deflectors. Therefore, lessons may be learned from this site to identify how flow type speed remains comparable to other similar sites despite the high levels of modification.

The three examples described above highlight how the comparison of a site to other sites with similar catchment-level effects may be beneficial for determining if the site in question is 'typical' of this type of river. Comparison of a site to pristine reference conditions is often preferable; for example, the WFD stipulates that the state of a waterbody should be compared to reference conditions of a stable ecosystem with an absence of human disturbance (Bouleau and Pont, 2015). However, there is a lack of such pristine conditions in highly modified landscapes as found in England (Macklin and Lewin, 2003). Also, setting goals for restoration based on reference conditions is problematic due to the weak distinction between natural variability in the system and the effects of human disturbance (Bouleau and Pont, 2015).

Here, similar sites are not considered as 'reference conditions' to restore back to a 'natural' state. The identification of similar sites based on catchment-level effects is instead designed to show the potential of sites for restoration based on sites being influenced by similar pressures. This approach adopts a move from an idealised view of restoration to the realities of river enhancement in degraded landscapes (Boon, 1992). Using the examples above, the flow habitats on the Wansbeck have the potential to be slightly faster by tackling the issue of poaching on the bank. The upland channel likely has a greater potential for natural recovery (Clarke *et al.*, 2003) so this would likely be a low-cost 'quick-win' for restoration. In comparison, the Ise has much lower flow type speeds than similar sites, likely due to channel resectioning. This represents a site that may require more extensive restoration measures such as the use of deflectors or bed raising to restore variations in depth and slow speed. The site on the Rase reflects a different use of this application, by locating a site that is typical of other similar reaches even with extensive modifications. Lessons could potentially be learnt from such a site to identify how it retains its faster, shallower flow types compared to other modified channels. Such simple comparisons in a national context allow for prioritisation of reaches for restoration based on an understanding of catchment-level effects.

6.5 CONCLUSIONS

This chapter provides a valuable insight into river functioning by combining characteristics that represent catchment-level effects at multiple spatial levels, including characteristics reflecting the channel network and its confluences that have not before been considered in such studies. The explanatory approach utilised was designed to explore the catchment-level effects, and the hierarchy of their interactions, that influence physical habitats rather than purely seeking predictive success (Mac Nally, 2000). This demonstrates the utility of exploring individual tree structure and interactions to explore these controls.

The modelling exercise identified a dominant upland-lowland gradient that can be captured by the waterbody typology developed in this thesis (Chapter 3). Within catchments a second gradient of upstream-downstream is present reflecting internal catchment functioning. These two high-level gradients are identified in previous studies (e.g. Jeffers, 1998; Vaughan *et al.*, 2013), but also reflect two key components of river system functioning described in Section 2.1.1. The novelty of this work stems from the inclusion of other features that describe low-level variation along these high-level gradients. The interactions between low-level characteristics and high-level characteristics shows that features of the network and tributaries, that were found to influence habitats in some cases (Chapters 4 and 5), increased the predictive power of the models and had a marked influence on habitat features under certain conditions.

It is also clear from the models presented in this chapter that habitat diversity has a more complex relationship with catchment-level effects than habitat type, reflecting the dynamism of the fluvial system at smaller spatial levels. The ability to explain even broad patterns in such variables highlights the utility of using broad-scale monitoring datasets for scientific enquiry, the overarching influence of catchment-level effects on river reaches, and the need for their inclusion in strategic management applications.

CHAPTER 7

Summary of findings, conclusions and future work

This thesis has explored the influence of catchment-level effects on physical habitats within reaches across rivers in England. This is because catchment controls on river reaches are of known importance but capturing the numerous catchment controls and their complex interactions is challenging. This means the true catchment-level effect is not included in large-scale studies (see evidence review, Section 2.1.3) or in river management. Therefore, an improved understanding of catchment-level effects may encourage more holistic integration of catchment-level effects in decision-making to aid more strategic management at a national level. This thesis has explored this issue by utilising a data-science approach; using information on physical habitats from numerous sites in combination with GIS datasets to explore associations between different catchment-level effects at multiple spatial levels. Here, the findings of each chapter are briefly summarised (Section 7.1) before the overall conclusions are discussed (Section 7.2). Opportunities for further development of the work conducted in this thesis is outlined in Section 7.3.

7.1 SUMMARY OF FINDINGS

The findings of this thesis are summarised below with respect to the specific objectives of each chapter, tabulated in **Table 2.4**. The number of each objective relates to the chapter within which the objective is addressed.

Objective 3a: *To build a typology of catchment-level effects that is practically useful for implementation by river managers.*

The work conducted in Chapter 3 met this objective by utilising a machine learning technique, self-organising maps (SOMs), to combine multiple natural and anthropogenic catchment characteristics into seven WFD waterbody types in England and Wales. The SOM method allowed the characteristics of each type to be interpreted visually through the production of numerous heatmaps (**Figure 3.1d**). These heatmaps highlighted correlations between variables, anomalies and categorical boundaries making SOM a useful method in comparison to classic ordination approaches (Astel *et al.*, 2007; Section 3.3.1). The seven types were organised primarily along an upland-lowland gradient, but also along a heterogeneity gradient from waterbodies with greater topographic roughness and dissection to more homogeneous waterbodies (see Section 3.2.4.3 for discussion). The typology is useful for river managers as it is developed using readily available GIS-derived data and allows continuous classification of catchment-level effects at a national level (see map of the typology classification in **Figure 3.2a**). It quantifies controls on reaches, rather than the response of river reaches, so is a complement to reach-level typologies.

Objective 3b: *To explore how effective a typology of catchment-level effects is at explaining physical habitats in river reaches.*

There were significant differences in physical habitat indices and compliance scores derived from the RHS dataset between the seven waterbody types (**Figure 3.3**). Physical habitats followed the upland-lowland and heterogeneity gradients identified from the SOM. There was variation in habitat indices within each waterbody type, despite statistical differences between those waterbody types. This internal variation was hypothesised to be a symptom of internal waterbody variation, such as upstream-downstream differences within waterbodies (**Figure 2.1**; Vannote *et al.*, 1980). Therefore, it was concluded that this typology would be useful for broad assessment of catchment-level effects and national management strategy, whereas reach-level management still requires more detailed knowledge of the site for decision-making.

Objective 4a: *To quantify network topology within catchments by creating a metric fit for multiple disciplinary use.*

The work in Chapter 4 repurposed two metrics from flood hydrograph estimation (network width function, Kirkby, 1976; link concentration function, Gupta *et al.*, 1986) to represent the density of links in a network over distance from the outlet and elevation above the outlet. The network density metrics captured the width of the network in addition to the longitudinal dimension of the network. The longitudinal dimension is usually the sole component of the network of interest, commonly represented by stream order (**Figure 5.1**). Network density metrics are easily extracted from a DEM and can be computed for any catchment. Chapter 4 tested the effect of the two network density metrics on physical habitats in four catchments in England. The results showed that network density influenced physical habitats in certain reaches but did not affect overall habitat condition in an area (**Figure 4.3**). Network density outperformed stream order in explaining habitat condition. There were also different responses observed between the four catchments so further work to identify why certain reaches and catchments are affected is conducted in Chapter 5.

Objective 5a: *To identify how important confluences influence the effect of network topology on river habitats.*

The network density metrics from Chapter 4 were calculated for 348 networks in England. Areas of high network density were expected to improve habitat condition, as shown in some catchments in Chapter 4, due to positive impact from numerous confluences (Benda *et al.*, 2004b). However, not all confluences in the network have an effect (Rice, 1998) which may explain why only 8-14% networks showed significant correlations between network density and habitat indices. The effect of confluences was assessed by comparing an RHS site upstream of the confluence to a site downstream for 853 confluences in England. Between 23% and 80% of confluences had an important effect on habitats depending on the habitat index (**Figure 5.9**) which is more than previously documented (Jones and Schmidt, 2016; Rice, 1998). The effect of confluences on habitat indices helps explain how network density is associated with habitats. For example, when a confluence which improved habitat condition was in a high-density area of the network, network density was positively related to overall habitat improvement in the area (**Figure 5.10**). However, in some cases local habitat improvements at the confluence did not influence overall habitat improvement in the area supporting the results of Rice *et al.* (2001). These results showed that both the effect of the confluence and the density of the network influenced the impact of network topology on physical habitat in rivers in England.

Objective 5b: *To investigate which properties of upstream tributaries influence confluence importance.*

The second objective of Chapter 5 was met by comparing the effect of confluences on habitats with the relative properties of the incoming tributaries. Previous research suggested that tributaries of similar size, but dissimilar properties would cause confluences to have greater effect on habitats downstream (see Section 5.1.1 and 5.1.2 for literature). Relatively weak correlations were identified between tributary properties and confluence effect on habitats (**Table 5.2**) which suggested that tributaries similar in size but dissimilar in other properties had the greatest impact on confluence effect, consistent with previous work. However, tributary dissimilarity did not always induce a positive effect on habitats, contrary to Kiffney *et al.* (2006). Results showed a tributary with anthropogenic land cover could worsen habitats in a comparatively natural stream, whereas natural land covers could improve habitats in highly modified streams. This analysis was the largest known study on confluence effects documented and highlights how in highly modified landscapes, such as England, confluences may reduce habitat quality rather than acting as a 'hotspot' of diversity as suggested by others (McClain *et al.*, 2003).

Objective 5c: *To explore how catchment morphometry influences the effect of network topology on river habitats.*

Most work on confluence effects and network structure has been conducted in high energy landscapes. One of the seminal theories suggests that more circular catchments will have more confluences that affect downstream reach characteristics (Benda *et al.*, 2004b). Therefore, the morphometry (elevation and circularity) of the catchment was compared to the effect of network density on habitat indices to understand why, in some catchments, habitats were influenced by network topology while others were not. Associations between morphometry and network density effect were relatively weak, but positive effects of network density on habitats were found in more circular and higher energy catchments (**Figure 5.11**) as expected based on previous findings (Benda *et al.*, 2004b; Ferguson *et al.*, 2006; Rice, 2017). These results showed that the catchment morphometry may influence the effect of network topology on river habitats.

Objective 6a: *To use a range of GIS-derived catchment and river properties at different spatial levels, to explain patterns of physical river habitats in England.*

The final objective addressed in Chapter 6 combined the waterbody types from Chapter 3, the network density metrics from Chapter 4 and the relative tributary properties from Chapter 5, along with upstream catchment properties and reach level variables that are commonly used in large-scale studies (see evidence review in Section 2.1.3). This chapter acts as a synthesis to the thesis by exploring how these multi-level variables interact to produce catchment-level effects on physical habitats. This objective was achieved using regression tree models and boosted regression trees (BRTs) to explore the hierarchical structure of interactions between the multi-level variables. The results showed that most variation in physical habitats is explained firstly by the type of waterbody a reach is situated in, and secondly by the reach's position in the catchment. Other features including network density and relative tributary properties offered further explanatory power under certain conditions highlighting that there are multiple levels of influence and interaction within catchment-level effects. The models produced were applied so simple comparisons could be made to identify whether an RHS site was 'typical' of other sites with similar catchment-level effects in order to prioritise potential sites for restoration.

7.2 OVERALL CONCLUSIONS

Three broad aims were introduced in Chapter 1 that reflect the principal foci of this thesis, the methods used to address it and the implications for river management. The diagram representing the interconnectivity between the aims is shown again in **Figure 7.1** and concluding remarks with regards to each aim are provided in the sections below.

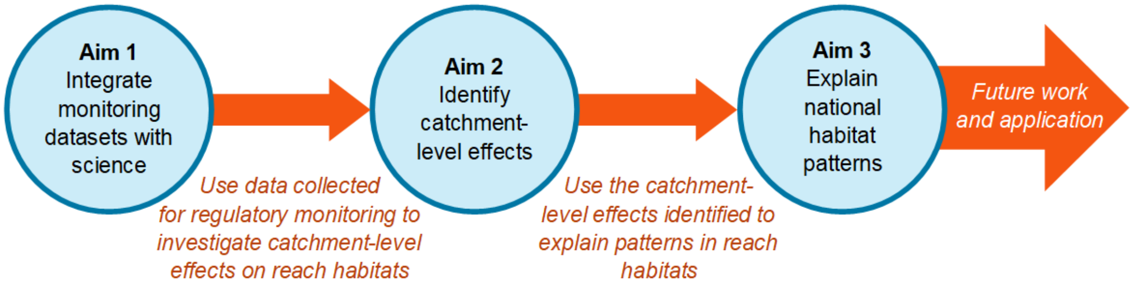


***Figure 7.1.** (Reworked **Figure 1.1**) Diagram showing the flow of the thesis aims towards future work and management applications. Connections between the aims described below each arrow.*

7.2.1 Monitoring data can answer scientific questions

An aim of this thesis was to explore catchment-level effects on river habitats nationally. The identification of national patterns is important because many studies focus on a small number of sites due to time and resource limitations, but contemporary studies must not only aim to expand knowledge, but also find methods of transferring knowledge to many, increasingly altered, catchments (Clifford, 2002). Therefore, the limitations of using broad-scale monitoring data collected by others for alternate purposes (Wessels *et al.*, 1998; Section 2.2.1) were deemed acceptable when the reward was numerous data points across a range of contrasting catchments.

The broad-scale RHS dataset had been used in the past for scientific enquiry (e.g. Jeffers, 1998; Harvey *et al.*, 2008; Vaughan *et al.*, 2013; Naura *et al.*, 2016), and was successfully used in this thesis in combination with data-science techniques (including correlation, PCA, SOM and BRT) to conduct analysis at the national level so that catchments, networks and reaches with different properties could be captured. This was critical to meet the aim of the thesis so that comparisons of sites within catchments and between different types of catchment could be made. Collecting such data first-hand would not have been possible within the time constraints of this PhD.

The broad-scale approach enabled patterns of catchment-level effects to be identified (Chapter 5 and 6) and tools built (Chapter 3 and 4) for England. It also allowed for the largest known study of confluence effects to be conducted (Chapter 5). The relatively low associations between different catchment-level effects and habitats within the reach partially reflect variation in sites and complexity of river processes that are captured by the RHS dataset. For example, associations with average measures of habitat type that adjusted to dominant reach conditions (Bunn and Arthington, 2002) were found to be stronger than the dynamic nature of habitat diversity (Padmore, 1997) using this data-science method on broad-scale data. The ability to derive broad patterns from such varied data adds credence to the associations identified.

7.2.2 Catchment-level effects explain patterns of physical habitats

Catchment-level effects are nebulous and this thesis both simplifies and adds further complexity to our understanding. In terms of simplifying the catchment-level effect, methods of data reduction including SOMs (Chapter 3) and PCA (Chapter 6) identify a gradient in catchment-level effects from upland to lowland which structure broad patterns of river habitats. This gradient controls the boundary conditions of catchments (Schumm and Lichty, 1965) which in turn determine the different processes occurring in different catchments.

Currently, a catchment's geology and land cover are frequently stated in studies but research on how findings may be extrapolated to different conditions are often lacking. This can lead to 'myths' about how river systems function, derived from few studies at small spatial scales, becoming guiding principles in river management without a holistic understanding of processes under different conditions (Calder and Aylward, 2006; Newson, 2010). The results from this thesis demonstrates the need for the inclusion of wider catchment characteristics in river research to contextualise results, and in management to constrain the outcomes of management activities (Beechie *et al.*, 2010) within these boundary conditions.

Along the upland-lowland gradient there is variation in habitats (**Figure 3.3**). In this thesis, the river network is considered an integrator of catchment-level effects, and variation within catchments is often explained by changes occurring from upstream to downstream along the river network (Jeffers, 1998a; Vaughan *et al.*, 2013). However, significant downstream trends in physical habitats were not present in >80% catchments in England (**Figure 5.7**), which Chapter 4 argues is because not all catchments exhibit a gradual change in habitats from upstream to downstream (**Figure 4.3**). Despite these findings, the upstream to downstream gradient is frequently used to describe the changing processes downstream (**Figure 2.2**; Vannote *et al.*, 1980; Church, 2002) and had strong associations with habitats when combined with the upland-lowland gradient (Section 6.4.1).

Yet both these broad trends within and between catchments do not capture the complexities of habitat variation present in England. Network topology was found to explain some of this variation (Chapter 4 and 5) and was the focus of this thesis. It is mechanistically important component of the catchment, connecting influences from up-catchment to the downstream channel (Tetzlaff *et al.*, 2007) and causing discontinuities in the upstream-downstream gradient (Rice *et al.*, 2001) but is overlooked by most previous large-scale studies (see evidence review in Section 2.1.3). Network topology was shown to have a weak overall influence on habitats compared to other factors, but it is still influential and important under certain catchment conditions (Chapter 6).

The recurring theme identified in this thesis is that while there are broad trends in habitats at the national level, which may be used to simplify our understanding of catchment-level effects, there are additional sources of complexity that have not been fully accounted for. One source of complexity, network topology, was simplified into metrics that could easily be applied in any large-scale study. This was a useful step in capturing some of this additional complexity, but there is more work to be done to isolate the influence of the network and identify under which conditions it is important.

7.2.3 Combine regional and upstream thinking for improved reach-level management

Often river management is conducted at the reach level, and while catchment-level effects are known to be important (Gilvear *et al.*, 2012), they are not often be considered fully due to limited project funding. The location of restoration sites is particularly opportunistic (Smith *et al.*, 2014), influenced by sympathetic land owners or funding constraints, rather than locating sites where restoration may have maximum effect based on understanding catchment-level effects. The results of this thesis highlight the utility of not only upstream thinking but regional thinking, i.e. considering what upstream characteristics may influence a reach of interest and placing the reach within the wider national context of catchment-level effects (Section 6.4.4). Therefore, lessons learnt from the management of a reach may be applied and adapted to other regions based on catchment-level effects.

Results also highlighted the importance of network topology, particularly how the density of the network and type of confluences may diminish or enhance anthropogenic effects from upstream (Chapter 5). From conversations with practitioners I have found that this phenomenon is not considered when designing restoration projects. However, it would be beneficial to conserve tributaries that are improving habitat conditions downstream and restore tributaries that are worsening habitat conditions to preserve the natural dynamism confluences add to the river system. This appreciation of network structure could be used to justify restoration works as benefits could be measured downstream rather just in the reach in question.

The results of this thesis were targeted at river managers via the datasets used to identify catchment-level effects (e.g. waterbody boundaries used for WFD compliance and other open access datasets) and using habitat variables extracted from the RHS dataset that are meaningful to river managers. Also, the development of the typology in Chapter 3 was designed to capture enough variation in catchment-level effects within waterbodies to be practical and useful to river managers, and the network density metrics in Chapter 4 were intended to be simply extracted for any catchment with a dendritic network so the method could easily be adopted by others. The multi-level tree model from Chapter 6 is also applied to a river management application. These tools and the knowledge gained from the analysis may help inform management practices and encourage a more strategic approach to catchment and river management.

7.2.4 Wider significance and future contributions of the research

This thesis is aligned with and can help deliver current catchment and river management approaches in the UK. Approaches such as upstream thinking, catchment visioning, nature conservation and designations, assessment of river condition and nature-based solutions are discussed below in relation to specific methods and results developed in the thesis.

The waterbody typology developed in Chapter 3 could be used to identify the range of waterbody types within a catchment of interest. This approach feeds into catchment-based and upstream thinking management approaches, where an initial step in the workflow is to use data and evidence to understand the baseline condition of the catchment and identify opportunities for improvements to river condition (CaBA, no date). The waterbody typology uses a range of catchment characteristics not usually considered by baseline assessments for a more holistic baseline including natural and anthropogenic controls on river quality.

The application of the multi-level model in Chapter 6 (**Figure 6.11**) demonstrates how utilising data from all the catchment, network, and reach levels can be used to identify sites where river habitat quality is worse than sites with similar controls. The automatic identification of such sites may supplement the catchment-based management approach by finding potential opportunities for river habitat improvement. This wider understanding of the multi-level controls on river habitat quality may also be beneficial for assessing river condition. In England, a newly mandated approach aims to achieve a net gain in biodiversity for all development projects. Assessments of river condition form one component of the Biodiversity Metric 2.0 (Crosher *et al.*, 2019) which measures biodiversity before and after development. The river condition assessment method builds on the pre-existing MoRPh survey (Gurnell *et al.*, 2018, 2019; Shuker *et al.*, 2017) to undertake field observations of physical habitats and features. These are combined with morphological assessments of longer subreaches and river condition is then evaluated within the context of the reach-scale geomorphological river type (Gurnell *et al.*, in review). The multi-level approach used in this thesis enables users of the Biodiversity Metric 2.0 to be realistic in setting post-development values by comparing the site in question to other sites with similar controls to estimate how much improvement is feasible.

While the multi-level model considers a range of different elements of the river system, there is also utility in focusing on the network density metrics developed in Chapter 4 alone. The range of habitats created at confluences and the presence of a greater number of tributaries in high density areas of the network indicate that these areas possess more refugia for aquatic biota during high flows (Koizumi *et al.*, 2013). Therefore, high areas of network density could be considered for protection or designation as zones of catchment resilience in the face of a changing climate where high flows are increasing in many 'natural' rivers in England (Harrigan *et al.*, 2018). The network density metrics also provide an approximate hydrograph for the catchment outlet as they were adapted from flood estimation methods. High network density areas therefore indicate the sections of the river network that may contribute to flood synchronisation. Natural Flood Management measures (nature-based solutions such as tree planting, implementing woody dams, restoring meanders etc.; Environment Agency, 2018b) may therefore be targeted in network dense areas to slow flows and help de-synchronise the flood peak.

The spatial nature of all components developed in this thesis mean that they may add additional dimensions to catchment visioning which displays the catchment and its features visually for improved communication with stakeholders (Taigel, 2016). Furthermore, the national extent of the metrics and datasets developed means that they may be applied in multiple catchments for streamlined and consistent assessment of catchment-level effects on river habitats.

7.3 OPPORTUNITIES FOR FUTURE WORK

Catchment-level effects on physical habitats in river reaches are complex. This thesis has improved understanding of some of these effects but several opportunities for future work are proposed:

(1) There are known limitations to the RHS database, with two in particular affecting the physical habitat indices used in this thesis (**Table 2.2**): (i) surveys were conducted within a 500m reach so different numbers of pool-riffle sequences are captured depending on river size (Emery *et al.*, 2004) which systematically influenced the habitat indices and; (ii) it was not possible to accurately calculate total habitat diversity as only dominant habitats are recorded at spot-checks, leaving marginal habitats such as backwaters unrecorded. While this thesis was still able to explain broad patterns in river habitat distribution across England, a useful development of this work would be to repeat some of this analysis with the MoRPh dataset (Shuker *et al.*, 2017) which has been growing in size and spatial coverage over the past three years. MoRPh varies survey length with river size to avoid the issues with one standard survey length and records all observed habitats for more accurate measures of diversity. It would be interesting to see whether the patterns observed in this thesis hold true, and whether the predictive capacity of the multi-level model improves or declines with more accurate data.

(2) Further work to isolate the influence on network topology would be a useful development. This thesis has shown how the network could be included in large-scale studies but future studies might explore how the density of important tributaries (rather than all tributaries as explored in Chapter 4) influences habitats, building on the observations made in Chapter 5 and the work of the Network Dynamic Hypothesis (Benda *et al.*, 2004b), and how this may vary between catchments with different properties.

(3) A hypothesis during the development of Chapter 5 was that confluences could be hotspots of catchment-level effects, reflecting characteristics in the upstream tributaries. This effect has been observed at individual confluences (**Figure 5.4b**) but no strong associations between tributary characteristics and confluence effects were identified using the national dataset in Chapter 5. Further work should identify whether the conditions at a confluence could be symptomatic of the processes occurring in their upstream tributaries by exploring additional variables and non-linear relationships between tributary properties and confluence effects. This would be useful to identify both which characteristics influence confluence effect and monitoring sites to detect changes upstream.

(4) This work did not consider the temporal dimension of catchment-level effects to instead focus on spatial patterns of river habitats. However, key drivers such as climate and land cover are likely to change over decadal time scales (Gurnell *et al.*, 2016) with growing pressures of climate change and population growth in the UK (Environment Agency, 2018a). It would be interesting to explore how predicted changes in these characteristics would impact catchment-level effects on physical habitats at a national level in England in order to target resources at high-risk areas.

Overall, the work presented in this thesis provides a novel contribution to river science by exploring key processes in the fluvial system in new ways using broad-scale data. This research was conducted with the backdrop of application to river management and this and future research should contribute to providing a holistic understanding catchment-level effects on river reaches for truly integrated river management.

References

Acreman, M., Dunbar, M., Hannaford, J., Mountford, O., Wood, P., Holmes, N., *et al.* (2008) Developing environmental standards for abstractions from UK rivers to implement the EU Water Framework Directive, *Hydrological Sciences Journal*, 53 (6), pp. 1105–1120. DOI:10.1623/hysj.53.6.1105.

Allan, J. D. (2004) Landcapes and Riverscapes: The influence of land use on stream ecosystems, *Annual Review of Ecology*, 35 (2004), pp. 257–284. DOI:10.1146/annurev.ecolsys.35.120202.110122.

ASCE (2000) Artificial Neural Networks in hydrology: preliminary concepts, *Journal of Hydrologic Engineering*, 5 (2), pp. 115–123. DOI:10.1061/(ASCE)1084-0699(2000)5:2(115).

Astel, A., Tsakovski, S., Barbieri, P. and Simeonov, V. (2007) Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets, *Water Research*, 41, pp. 4566–4578. DOI:10.1016/j.watres.2007.06.030.

Atkinson, C. L., Julian, J. P. and Vaughn, C. C. (2012) Scale-dependent longitudinal patterns in mussel communities, *Freshwater Biology*, 57 (11), pp. 2272–2284. DOI:10.1111/fwb.12001.

Azam, M., Park, H. K., Maeng, S. J. and Kim, H. S. (2017) Regionalization of drought across South Korea using multivariate methods, *Water*, 10 (1), pp. 1–23. DOI:10.3390/w10010024.

Baartman, J. E. M., Masselink, R., Keesstra, S. D. and Temme, A. J. A. M. (2013) Linking landscape morphological complexity and sediment connectivity, *Earth Surface Processes and Landforms*, 38 (12), pp. 1457–1471. DOI:10.1002/esp.3434.

Bannister, N., Mant, J. and Janes, M. (2005) A review of catchment scale river restoration projects in the UK, *The River Restoration Centre*, pp. 1–42.

Beechie, T. J., Sear, D. A., Olden, J. D., Pess, G. R., Buffington, J. M., Moir, H., Roni, P. and Pollock, M. M. (2010) Process-based principles for restoring river ecosystems, *BioScience*, 60 (3), pp. 209–222. DOI:10.1525/bio.2010.60.3.7.

Belletti, B., Rinaldi, M., Buijse, A. D., Gurnell, A. M. and Mosselman, E. (2015) A review of assessment methods for river hydromorphology, *Environmental Earth Sciences*, 73, pp. 2079–2100. DOI:10.1007/s12665-014-3558-1.

Belsky, A. J., Matzke, A. and Uselman, S. (1999) Survey of livestock influences on stream and riparian ecosystems in the western United States, *Journal of Soil and Water Conservation*, 54 (1), pp. 419–431.

Benda, L., Andras, K., Miller, D. and Bigelow, P. (2004a) Confluence effects in rivers: interactions of basin scale, network geometry, and disturbance regimes, *Water Resources Research*, 40 (5), pp. 1–15. DOI:10.1029/2003WR002583.

Benda, L., Poff, N. L., Miller, D., Dunne, T., Reeves, G., Pess, G. and Pollock, M. (2004b) The network dynamics hypothesis: how channel networks structure riverine habitats, *BioScience*, 54 (5), pp. 413–427. DOI:10.1641/0006-3568(2004)054[0413:TNDHHC]2.0.CO;2.

Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1), pp. 289–300. DOI:10.1111/j.2517-6161.1995.tb02031.x.

Benone, N. L., Esposito, M. C., Juen, L., Pompeu, P. S. and Montag, L. F. A. (2017) Regional controls on physical habitat structure of Amazon streams, *River Research and Applications*, 33 (5), pp. 766–776. DOI:10.1002/rra.3137.

Berrie, A. D. (1992) The chalk-stream environment, *Hydrobiologia*, 248 (1), pp. 3–9. DOI:10.1007/BF00008881.

Best, J. L. (1985) *Flow dynamics and sediment transport at river channel confluences*. Birbeck. University of London.

Best, J. L. (1986) The morphology of river channel confluences, *Progress in Physical Geography*, 10, pp. 157–174. DOI:10.1177/030913338601000201.

Best, J. L. (1987) Flow dynamics at river channel confluences: implications for sediment transport and bed morphology, *Recent Developments in Fluvial Sedimentology*, SP39, pp. 27–35. DOI:10.2110/pec.87.39.0027.

Beven, K. J. and Kirkby, M. J. (1979) A physically based, variable contributing area model of basin hydrology, *Hydrological Sciences Bulletin*, 24 (1), pp. 43–69. DOI:10.1080/02626667909491834.

BGS [no date] Bedrock Geology 1:625,000, *British Geological Society*. Available from: http://www.bgs.ac.uk/products/digitalmaps/DiGMapGB_50.html [Accessed 2 March 2016].

Biron, P. M., Richer, A., Kirkbride, A. D., Roy, A. G. and Han, S. (2002) Spatial patterns of water surface topography at a river confluence, *Earth Surface Processes and Landforms*, 27 (9), pp. 913–928. DOI:10.1002/esp.359.

Bizzi, S. and Lerner, D. N. (2012) Characterizing physical habitats in rivers using map-derived drivers of fluvial geomorphic processes, *Geomorphology*, 169–170, pp. 64–73. DOI:10.1016/j.geomorph.2012.04.009.

Bizzi, S. and Lerner, D. N. (2015) The use of stream power as an indicator of channel sensitivity to erosion and deposition processes, *River Research and Applications*, 31, pp. 16–27. DOI:10.1002/rra.

Blei, D. M. and Smyth, P. (2017) Science and data science, *Proceedings of the National Academy of Sciences of the United States of America*, 114 (33), pp. 8689–8692. DOI:10.1073/pnas.1702076114.

Boddy, N. C., Booker, D. J. and McIntosh, A. R. (2019) Confluence configuration of river networks controls spatial patterns in fish communities, *Landscape Ecology*, 34 (1), pp. 187–201. DOI:10.1007/s10980-018-0763-4.

Boon, P. J. (1992) Essential elements in the case for river conservation, in: Calow, P. and Petts, G. E. (eds.) *River Conservation and Management*. Chichester: John Wiley and Sons, pp. 11–34.

Bouleau, G. and Pont, D. (2015) Did you say reference conditions? Ecological and socio-economic perspectives on the European Water Framework Directive, *Environmental Science & Policy*, 47, pp. 32–41. DOI:10.1016/j.envsci.2014.10.012.

Brandon, G., Boehmke, B. and Cunningham, J. (2019) Generalized Boosted Regression Models, *Package 'gbm' version 2.1.5. Originally developed by Ridgeway*. Available from: https://cran.r-project.org/web/packages/gbm/gbm.pdf [Accessed 1 July 2019].

Bravard, J. P. and Gilvear, D. J. (1996) Hydrological and geomorphological structure of hydrosystems, in: Petts, G. E. and Amoros, C. (eds.) *The Fluvial Hydrosystem*. Springer Netherlands, pp. 98–116.

Brierley, G. and Fryirs, K. (2000) River Styles, a geomorphic approach to catchment characterization: implications for river rehabilitation in Bega Catchment, New South Wales, Australia., *Environmental Management*, 25 (6), pp. 661–679. DOI:10.1007/s002670010052.

Brookes, A. (1988) *Channelized rivers, perspectives for environmental management*. New York: John Wiley & Sons.

Buendia, C., Gibbins, C. N., Vericat, D. and Batalla, R. J. (2013) Reach and catchment-scale influences on invertebrate assemblages in a river with naturally high fine sediment loads, *Limnologica*, 43 (5), pp. 362–370. DOI:10.1016/j.limno.2013.04.005.

Bunn, S. E. and Arthington, A. H. (2002) Basic principles and ecological consequences of altered flow regimes for aquatic biodiversity, *Environmental Management*, 30 (4), pp. 492–507. DOI:10.1007/s00267-002-2737-0.

CaBA [no date] CaBA Workflow. Available from: https://catchmentbasedapproach.org/about/caba-workflow/

Calder, I. R. and Aylward, B. (2006) Forests and floods. Moving to an evidence-based approach to watershed and integrated flood management, *Water International*, 31 (1), pp. 87–99.

Cashman, M. J. (2014) *The effect of large wood on river physical habitat and nutritional dynamics*. Queen Mary, University of London, Thesis.

CEH [no date-a] 1:50,000 Digital River Network, *Centre for Ecology and Hydrology*. Available from:

https://www.ceh.ac.uk/services/150000-watercourse-network [Accessed 2 February 2016].

CEH [no date-b] Integrated Hydrological Digital Terrain Model, *Centre for Ecology and Hydrology*. Available from: https://www.ceh.ac.uk/services/integrated-hydrological-digital-terrain-model [Accessed 2 February 2016].

CEH, Morton, R. D., Rowland, C. S., Wood, C. M., Meek, L., Marston, C. G. and Smith, G. M. (2014) Land Cover Map 2007 (25m raster, GB) v1.2., *Centre for Ecology and Hydrology*. DOI:10.5285/a1f88807-4826-44bc-994d-a902da5119c2.

CEN (2004) Water quality - Guidance standard for assessing the hydromorphological features of rivers, *European Committee for Standardization*, EN 14614.

Cericola, F., Portis, E., Toppino, L., Barchi, L., Acciarri, N., Ciriaci, T., Sala, T., Rotino, G. L. and Lanteri, S. (2013) The population structure and diversity of eggplant from Asia and the Mediterranean Basin, *PLoS ONE*, 8 (9), pp. e73702. DOI:10.1371/journal.pone.0073702.

Chorley, R. J. (1969) The drainage basin as the fundamental geomorphic unit, in: Chorley, R. J. (ed.) *Water, earth, and man: a synthesis of hydrology, geomorphology and socio-economic geography*. London: Methuen, pp. 77–99.

Church, M. (2002) Geomorphic thresholds in riverine landscapes, *Freshwater Biology*, 47, pp. 541–557. DOI:10.1046/j.1365-2427.2002.00919.x.

Church, M. and Ferguson, R. I. (2015) Morphodynamics: rivers beyond steady state, *Water Resources Research*, 51, pp. 1883–1897. DOI:10.1002/2014WR016862.

Church, M. and Kellerhals, R. (1978) On the statistics of grain size variation along a gravel river, *Canadian Journal of Earth Sciences*, 15 (7), pp. 1151–1160.

Clarke, S. J., Bruce-Burgess, L. and Wharton, G. (2003) Linking form and function: towards an eco-hydromorphic approach to sustainable river restoration, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13 (5), pp. 439–450. DOI:10.1002/aqc.591.

Clifford, N. J. (2002) Hydrology: the changing paradigm, *Progress in Physical Geography*, 26 (2), pp. 290–301.

Clifford, N. J., Harmar, O. P., Harvey, G. and Petts, G. E. (2006) Physical habitat, eco-hydraulics and river design: a review and re-evaluation of some popular concepts and methods, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 16 (4), pp. 389–408. DOI:10.1002/aqc.736.

Cockburn, J. M. H., Villard, P. V and Hutton, C. (2015) Assessing instream habitat suitability and hydraulic signatures of geomorphic units in a reconstructed single thread meandering channel, *Ecohydrology*. DOI:10.1002/eco.1705.

Cohen, P., Andriamahefa, H. and Wasson, J.-G. (1998) Towards a regionalizaton of aquatic habitat: distribution of mesohabitats at the scale of a large basin., *Regulated Rivers: Research and Management*, 14 (1), pp. 391–404. DOI:10.1002/(SICI)1099-1646(199809/10)14:5<391::AID-RRR513>3.0.CO;2-W.

Collins, A., Miller, J., Coughlin, D. and Kirk, S. (2014) The production of Quick Scoping Reviews and Rapid Evidence Assessments: a how to guide, *Joint Water Evidence Group, Beta Version 2*.

Cooksley, S. L., Brewer, M. J., Donnelly, D., Spezia, L. and Tree, A. (2012) Impacts of artificial structures on the freshwater pearl mussel Margaritifera margaritifera in the River Dee, Scotland, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 22 (3), pp. 318–330. DOI:10.1002/aqc.2241.

Crosher, I., Gold, S., Heaver, M., Heydon, M., Moore, L., Panks, S., Scott, S., Stone, D. and White, N. (2019) The Biodiversity Metric 2.0: auditing and accounting for biodiversity value. User guide (Beta Version, July 2019)., *Natural England*. Available from: http://publications.naturalengland.org.uk/publication/5850908674228224 [Accessed 3 May 2020].

Dadson, S. J., Hall, J. W., Murgatroyd, A., Acreman, M., Bates, P., Beven, K., *et al.* (2017) A restatement of the natural science evidence concerning flood management in the UK, *Proceedings of the Royal Society A*, 473. DOI:10.1109/18.75242.

Davenport, A. J., Gurnell, A. M. and Armitage, P. D. (2004) Habitat survey and classification of urban rivers, *River Research and Applications*, 20 (6), pp. 687–704. DOI:10.1002/rra.785.

Davies, D. L. and Bouldin, D. W. (1979) A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Learning Intelligence*, 1 (2), pp. 224–227. DOI:10.1109/TPAMI.1979.4766909.

Davies, N. M., Norris, R. H. and Thoms, M. C. (2000) Prediction and assessment of local stream habitat features using large-scale catchment characteristics, *Freshwater Biology*, 45, pp. 343–369.

de Castro, D. M. P., Dolédec, S. and Callisto, M. (2017) Landscape variables influence taxonomic and trait composition of insect assemblages in Neotropical savanna streams, *Freshwater Biology*, 62 (8), pp. 1472–1486. DOI:10.1111/fwb.12961.

de Paula, F. R., Gerhard, P., Wenger, S. J., Ferreira, A., Vettorazzi, C. A. and de Barros Ferraz, S. F. (2013) Influence of forest cover on in-stream large wood in an agricultural landscape of southeastern Brazil: a multi-scale analysis, *Landscape Ecology*, 28 (1), pp. 13–27. DOI:10.1007/s10980-012-9809-1.

Death, R. G. and Joy, M. K. (2004) Invertebrate community structure in streams of the Manawatu-Wanganui region, New Zealand: the roles of catchment versus reach scale influences, *Freshwater Biology*, 49 (8), pp. 982–997. DOI:10.1111/j.1365-2427.2004.01243.x.

Debnath, J., Das (Pan), N., Sharma, R. and Ahmed, I. (2019) Impact of confluence on hydrological and morphological characters of the trunk stream: a study on the Manu River of North-east India, *Environmental Earth Sciences*, 78 (6), pp. 1–19. DOI:10.1007/s12665-019-8190-7.

Defra (2013) Catchment Based Approach: improving the quality of our water environment, *Department for Environment, Food & Rural Affairs*. Available from: https://www.gov.uk/government/publications/catchment-based-approach-improving-the-quality-of-our-water-environment

Depettris, C., Mendiondo, E. M., Neiff, J. and Rohrmann, H. (2000) Flood defence strategy at the confluence of the Paraná-Paraguay rivers, *Proc. Int. Symp. Flood Defence*, Kassel: He, pp. C31,C40.

Dixon, S. J., Sambrook Smith, G. H., Best, J. L., Nicholas, A. P., Bull, J. M., Vardy, M. E., Sarker, M. H. and Goodbred, S. (2018) The planform mobility of river channel confluences: insights from analysis of remotely sensed imagery, *Earth-Science Reviews*, 176, pp. 1–18. DOI:10.1016/j.earscirev.2017.09.009.

Dollar, E., James, C., Rogers, K. and Thoms, M. (2007) A framework for interdisciplinary understanding of rivers as ecosystems, *Geomorphology*, 89 (1–2), pp. 147–162.

Dovers, S. R. and Day, D. G. (1988) Australian rivers and statute law, *Environmental Planning and Law Journal*, 5, pp. 90–108.

Downs, P. W. and Gregory, K. J. (2004) *River channel management: towards sustainable catchment hydrosystems*, *River Channel Management*. New York: Routledge.

Downs, P. W., Gregory, K. J. and Brookes, A. (1991) How integrated is river basin management?, *Environmental Management*, 15 (3), pp. 299–309. DOI:10.1007/BF02393876.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., *et al.* (2006) Freshwater biodiversity: importance, threats, status and conservation challenges, *Biological Reviews*, 81 (2), pp. 163–182. DOI:10.1017/S1464793105006950.

Duncan, W. W., Poole, G. C. and Meyer, J. L. (2009) Large channel confluences influence geomorphic heterogeneity of a southeastern United States river, *Water Resources Research*, 45 (10), pp. 1–9. DOI:10.1029/2008WR007454.

Dupas, R., Delmas, M., Dorioz, J. M., Garnier, J., Moatar, F. and Gascuel-Odoux, C. (2015) Assessing the impact of agricultural pressures on N and P loads and eutrophication risk, *Ecological Indicators*, 48, pp. 396–407. DOI:10.1016/j.ecolind.2014.08.007.

Elith, J., Leathwick, J. R. and Hastie, T. (2008) A working guide to boosted regression trees, *Journal of Animal Ecology*, 77 (4), pp. 802–813. DOI:10.1111/j.1365-2656.2008.01390.x.

Emery, J. C., Gurnell, A. M., Clifford, N. J. and Petts, G. E. (2004) Characteristics and controls of gravel-bed riffles: An analysis of data from the river-habitat survey, *Water and Environment Journal*, 18 (4), pp. 210–216. DOI:10.1111/j.1747-6593.2004.tb00535.x.

England, J. and Gurnell, A. M. (2016) Incorporating catchment to reach scale processes into hydromorphological assessment in the UK, *Water and Environment Journal*, 30 (1–2), pp. 22–30.

DOI:10.1111/wej.12172.

Environment Agency (2003) *River Habitat Survey in Britain and Ireland field survey guidance manual: 2003 version*. Available from: https://www.gov.uk/government/uploads/system/.../LIT_1758.pdf

Environment Agency (2010) Our river habitats: River habitats in England and Wales: current state and changes since 1995-96, pp. 1–31. Available from: www.riverhabitatsurvey.org/manual/reports/

Environment Agency (2014) WFD Surface Water Management Catchments Cycle 2. Available from: https://data.gov.uk/dataset/1a494e3e-e414-456c-9c2e-ca367a2945b6/wfd-surface-water-management-catchments-cycle-2

Environment Agency (2018a) The state of the environment: water quality. Available from: https://www.gov.uk/government/publications/state-of-the-environment

Environment Agency (2018b) Working with Natural Processes – Evidence Directory. Available from: https://www.gov.uk/government/publications/working-with-natural-processes-to-reduce-flood-risk [Accessed 21 May 2020].

European Commission (2000) Directive 2000/60/EC, *Official Journal of the European Communities*, L 269, pp. 1–15.

Evans, I. S. and Minár, J. (2011) A classification of geomorphometric variables, in: *International Geomorphometry 2011*. Redlabds, CA, pp. 105–108.

Farley, S. S., Dawson, A., Goring, S. J. and Williams, J. W. (2018) Situating ecology as a big-data science: Current advances, challenges, and solutions, *BioScience*, 68 (8), pp. 563–576. DOI:10.1093/biosci/biy068.

Fausch, K. D., Torgersen, C. E., Baxter, C. V., Li, H. W., View, C., The, O. F., *et al.* (2002) Landscapes to riverscapes: bridging the gap between research and conservation of stream fishes, *BioScience*, 52 (6), pp. 483–498. DOI:10.1641/0006-3568(2002)052[0483:LTRBTG]2.0.CO;2.

Feld, C. K. (2004) Identification and measure of hydromorphological degradation in Central European lowland streams, *Hydrobiologia*, 516, pp. 69–90. DOI:10.1023/B.

Feld, C. K., Segurado, P. and Gutiérrez-Cánovas, C. (2016) Analysing the impact of multiple stressors in aquatic biomonitoring data: A ‘cookbook’ with applications in R, *Science of the Total Environment*, 573, pp. 1320–1339. DOI:10.1016/j.scitotenv.2016.06.243.

Ferguson, R. I., Cudden, J. R., Hoey, T. B. and Rice, S. P. (2006) River system discontinuities due to lateral inputs: generic styles and controls, *Earth Surface Processes and Landforms*, 31 (9), pp. 1149–1166. DOI:10.1002/esp.1309.

Ferguson, R. I., Hoey, T. B., Wathen, S. and Werritty, A. (1996) Field evidence for rapid downstream fining of river gravels through selective transport, *Geology*, 24 (2), pp. 179–182. DOI:10.1130/0091-7613(1996)024<0179:FEFRDF>2.3.CO;2.

Fernandes, C. C. (2004) Amazonian ecology: tributaries enhance the diversity of electric fishes, *Science*, 305 (5692), pp. 1960–1962. DOI:10.1126/science.1101240.

Flores, A. N., Bledsoe, B. P., Cuhaciyan, C. O. and Wohl, E. E. (2006) Channel-reach morphology dependence on energy, scale, and hydroclimatic processes with implications for prediction using geospatial data, *Water Resources Research*, 42, pp. 1–15. DOI:10.1029/2005WR004226.

Fox, P. J. A., Naura, M. and Scarlett, P. (1998) An account of the derivation and testing of a standard field method, River Habitat Survey, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 8 (4), pp. 455–475. DOI:10.1002/(SICI)1099-0755(199807/08)8:4<455::AID-AQC284>3.0.CO;2-7.

Frappier, B. and Eckert, R. T. (2007) A new index of habitat alteration and a comparison of approaches to predict stream habitat conditions, *Freshwater Biology*, 52 (10), pp. 2009–2020. DOI:10.1111/j.1365-2427.2007.01803.x.

Friedman, J. H. (2001) Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 29, pp. 1189–1232.

Frissell, C. A., Liss, W. J., Warren, C. E. and Hurley, M. D. (1986) A hierarchical framework stream habitat classification, *Environmental Management*, 10 (2), pp. 199–214. DOI:10.1007/BF01867358.

Fröhlich, H. L., Breuer, L., Vaché, K. B. and Frede, H. G. (2008) Inferring the effect of catchment complexity on mesoscale hydrologic response, *Water Resources Research*, 44 (W09414), pp. 1–15. DOI:10.1029/2007WR006207.

Gardner, R. H., Milne, B. T., Turnei, M. G. and O'Neill, R. V. (1987) Neutral models for the analysis of broad-scale landscape pattern, *Landscape Ecology*, 1 (1), pp. 19–28. DOI:10.1007/BF02275262.

Gilvear, D. J., Casas-Mulet, R. and Spray, C. J. (2012) Trends and issues in delivery of integrated catchment scale river restoration., *River Research and Applications*, 28, pp. 234–246. DOI:10.1002/rra.

Gilvear, D. J., Spray, C. J. and Casas-Mulet, R. (2013) River rehabilitation for the delivery of multiple ecosystem services at the river network scale, *Journal of Environmental Management*, 126, pp. 30–43. DOI:10.1016/j.jenvman.2013.03.026.

Gilvear, D. J. and Willby, N. J. (2006) Channel dynamics and geomorphic variability as controls on gravel bar vegetation; River Tummel, Scotland, *River Research and Applications*, 22 (4), pp. 457–474. DOI:10.1002/rra.917.

Glickman, M. E., Rao, S. R. and Schultz, M. R. (2014) False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies, *Journal of Clinical Epidemiology*, 67 (8), pp. 850–857. DOI:http://dx.doi.org/10.1016/j.jclinepi.2014.03.012.

Gostner, W., Alp, M., Schleiss, A. J. and Robinson, C. T. (2013) The hydro-morphological index of diversity: A tool for describing habitat heterogeneity in river engineering projects, *Hydrobiologia*, 712 (1), pp. 43–60. DOI:10.1007/s10750-012-1288-5.

Gregory, K. J. and Walling, D. E. (1973) *Drainage basin form and process*. Chichester: Wiley.

Gupta, V. K. and Mesa, O. J. (1988) Runoff generation and hydrologic response via channel network geomorphology - recent progress and open problems, *Journal of Hydrology*, 102 (1–4), pp. 3–28. DOI:10.1016/0022-1694(88)90089-3.

Gupta, V. K. and Waymire, E. D. (1983) On the formulation of an analytical approach to hydrologic response and similarity at the basin scale, *Journal of Hydrology*, 65 (1–3), pp. 95–123.

Gupta, V. K., Waymire, E. D. and Rodriguez-Iturbe, I. (1986) On scales, gravity and network structure in basin runoff, in: *Scale problems in hydrology*. Springer Netherlands, pp. 159–184.

Gurnell, A. M., England, J., Shuker, L. and Wharton, G. (2019) The contribution of citizen science volunteers to river monitoring and management: international and national perspectives and the example of the MoRPh survey, *River Research and Applications*, 35 (8), pp. 1359–1373. DOI:10.1002/rra.3483.

Gurnell, A. M., O'Hare, J. M., O'Hare, M. T., Dunbar, M. J. and Scarlett, P. M. (2010) An exploration of associations between assemblages of aquatic plant morphotypes and channel geomorphological properties within British rivers, *Geomorphology*, 116, pp. 135–144. DOI:10.1016/j.geomorph.2009.10.014.

Gurnell, A. M., Rinaldi, M., Belletti, B., Bizzi, S., Blamauer, B., Braca, G., *et al.* (2016) A multi-scale hierarchical framework for developing understanding of river behaviour to support river management, *Aquatic Sciences*, 78 (1), pp. 1–16. DOI:10.1007/s00027-015-0424-5.

Gurnell, A. M., Scott, S. J., England, J., Gurnell, D. J., Jeffries, R., Shuker, L. and Wharton, G. (in review) Assessing river condition: a multiscale approach designed for operational application in the context of Biodiversity Net Gain, *River Research and Applications*.

Gurnell, A., Shuker, L., Wharton, G. and England, J. (2018) The MoRPh Survey: A Modular River Physical Survey for Citizen Scientists, *Technical Reference Manual*, pp. 1–42.

Gustafson, E. J. (1998) Quantifying landscape spatial pattern: what is the state of the art?, *Ecosystems*, 1, pp. 143–156. DOI:10.1007/s100219900011.

Harper, D. M. and Everard, M. (1998) Why should the habitat-level approach underpin holistic river survey and management ?, 413, pp. 395–413.

Harper, D. M., Smith, C., Barham, P. J. and Howell, R. (1995) The ecological basis for the management of the natural river environment, in: Harper, D. M. and Ferguson, A. J. . (eds.) *The ecological basis for river management*. Chichester: Wiley, pp. 219–238.

Harper, D., Smith, C. and Barham, P. (1992) Habitats as the building blocks for river conservation assessment, in: Boon, P. J., Calow, P., and Petts, G. E. (eds.) *River conservation and management*. New York: Wiley, pp. 311–319.

Harrigan, S., Hannaford, J., Muchan, K. and Marsh, T. J. (2018) Designation and trend analysis of the updated UK Benchmark Network of river flow stations: The UKBN2 dataset, *Hydrology Research*, 49 (2), pp. 552–567. DOI:10.2166/nh.2017.058.

Harvey, G. L. and Clifford, N. J. (2008) Distribution of biologically functional habitats within a lowland river, United Kingdom, *Aquatic Ecosystem Health and Management Management*, 11 (January), pp. 465–473. DOI:10.1080/14634980802515682.

Harvey, G. L., Clifford, N. J. and Gurnell, A. M. (2008a) Towards an ecologically meaningful classification of the flow biotope for river inventory, rehabilitation, design and appraisal purposes, *Journal of Environmental Management*, 88, pp. 638–650. DOI:10.1016/j.jenvman.2007.03.039.

Harvey, G. L., Gurnell, A. M. and Clifford, N. J. (2008b) Characterisation of river reaches: the influence of rock type, *Catena*, 76 (1), pp. 78–88. DOI:10.1016/j.catena.2008.09.010.

Harvey, G. L. and Wallerstein, N. P. (2009) Exploring the interactions between flood defence maintenance works and river habitats: the use of River Habitat Survey data, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 19, pp. 689–702.

Hastie, T., Tibshirani, R. and Friedman, J. (2001) *The elements of statistical learning: data mining, inference and prediction*. New York: Springer.

Heasley, E. L., Clifford, N. J. and Millington, J. D. A. (2019) Integrating network topology metrics into studies of catchment-level effects on river characteristics, *Hydrology and Earth System Sciences*, 23 (5), pp. 2305–2319. DOI:10.5194/hess-23-2305-2019.

Heasley, E. L., Millington, J. D. A., Clifford, N. J. and Chadwick, M. A. (2020) A waterbody typology derived from catchment controls using self-organising maps, *Water*, 12 (1), pp. 1–20. DOI:10.3390/w12010078.

Helsel, B. D. R. and Hirsch, R. M. (2002) Statistical methods in water resources, in: *Book 4, Hydrologic Analysis and Interpretation: Techniques of Water-Resources Investigations*. Reston, VA: United States Geological Survey,.

Hill, G., Maddock, I. and Bickerton, M. (2008) River habitat mapping: are surface flow type habitats biologically distinct?, *BHS 10th National Hydrology Symposium*, pp. 165–171.

Holmes, M. G. R., Young, A. R., Gustard, A. and Grew, R. (2002) A region of influence approach to predicting flow duration curves within unguaged catchments, *Hydrology and Earth System Sciences*, 6 (4), pp. 721–731. DOI:10.5194/hess-6-721-2002.

Holmes, N. T. H., Boon, P. J. and Rowell, T. A. (1998) A revised classification system for British rivers based on their aquatic plant communities, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 8, pp. 555–578. DOI:10.1002/(SICI)1099-0755(199807/08)8:4<555::AID-AQC296>3.0.CO;2-Y.

Hopkins, R. L. (2009) Use of landscape pattern metrics and multiscale data in aquatic species distribution models: a case study of a freshwater mussel, *Landscape Ecology*, 24 (7), pp. 943–955. DOI:10.1007/s10980-009-9373-5.

Hornby, D. D. (2010) RivEX. version 6.7, *http://www.rivex.co.uk*. Available from: http://www.rivex.co.uk.

Husson, F., Josse, J., Le, S. and Mazet, J. (2018) Multivariate exploratory data analysis and data mining, *Package 'FactoMineR" v1.41'*. DOI:10.1201/b10345-2>.License.

Husson, F., Josse, J. and Pages, J. (2010) Principal component methods, hierarchical clustering and partitional clustering: why would we need to choose for visualizing data?, *Technical Report - Agrocampus, Applied Mathematics Department*.

Hutchens, J. J., Schuldt, J. A., Richards, C., Johnson, L. B., Host, G. E. and Breneman, D. H. (2009) Multi-scale mechanistic indicators of Midwestern USA stream macroinvertebrates, *Ecological Indicators*, 9 (6), pp. 1138–1150. DOI:10.1016/j.ecolind.2009.01.001.

Hynes, H. B. N. (1975) The stream and its valley, *Edgardo Baldi Memorial Lecture, SIL Proceedings*, 19 (1), pp. 1–15. DOI:10.1080/03680770.1974.11896033.

Jacobson, R. B., Femmer, S. R. and McKenney, R. A. (2001) *Land-use changes and the physical habitat of streams: a review with emphasis on studies within the US Geological Survey Federal State Cooperative Program*. 1175th ed. Reston, VA: US Geological Survey. DOI:10.3133/cir1175.

Jähnig, S. C., Shah, D. N., Tachamo Shah, R. D., Li, F., Cai, Q., Sundermann, A., Tonkin, J. D. and Stendera, S. (2015) Community–environment relationships of riverine invertebrate communities in central Chinese streams, *Environmental Earth Sciences*, 74 (8), pp. 6431–6442. DOI:10.1007/s12665-015-4466-8.

Jeffers, J. N. R. (1998a) Characterization of river habitats and prediction of habitat features using ordination techniques, *Aquatic Conservation: Marine and Freshwater Ecosystems (B)*, 8 (4), pp. 529–540. DOI:10.1002/(SICI)1099-0755(199807/08)8:4<529::AID-AQC301>3.0.CO;2-9.

Jeffers, J. N. R. (1998b) The statistical basis of sampling strategies for rivers: An example using River Habitat Survey, *Aquatic Conservation: Marine and Freshwater Ecosystems (A)*, 8 (4), pp. 447–454. DOI:10.1002/(SICI)1099-0755(199807/08)8:4<447::AID-AQC288>3.0.CO;2-R.

Jensen, C. K., McGuire, K. J., Shao, Y. and Andrew Dolloff, C. (2018) Modeling wet headwater stream networks across multiple flow conditions in the Appalachian Highlands, *Earth Surface Processes and Landforms*, 43 (13), pp. 2762–2778. DOI:10.1002/esp.4431.

Jones, N. E. and Schmidt, B. J. (2016) Tributary effects in rivers: interactions of spatial scale, network structure, and landscape characteristics, *Canadian Journal of Fisheries and Aquatic Sciences*, 74 (4), pp. 503–510. DOI:10.1139/cjfas-2015-0493.

Jones, N. E. and Schmidt, B. J. (2018) Influence of tributaries on the longitudinal patterns of benthic invertebrate communities, *River Research and Applications*, 34 (2), pp. 165–173. DOI:10.1002/rra.3240.

Jowett, I. G. (1993) A method for objectively identifying pool, run, and riffle habitats from physical measurements, *New Zealand Journal of Marine and Freshwater Research*, 27 (2), pp. 241–248. DOI:10.1080/00288330.1993.9516563.

Jusik, S., Szoszkiewicz, K., Kupiec, J. M., Lewin, I. and Samecka-Cymerman, A. (2015) Development of comprehensive river typology based on macrophytes in the mountain-lowland gradient of different Central European ecoregions, *Hydrobiologia*, 745 (1), pp. 241–262. DOI:10.1007/s10750-014-2111-2.

Karaouzas, I., Gritzalis, K. C. and Skoulikidis, N. (2007) Land use effects on macroinvertebrate assemblages and stream quality along an agricultural river basin, *Fresenius Environmental Bulletin*, 16 (6), pp. 645–653.

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N. and Kumar, V. (2017) Theory-guided data science: a new paradigm for scientific discovery from data, *IEEE Transactions on Knowledge and Data Engineering*, 29 (10), pp. 2318–2331. DOI:10.1109/tkde.2017.2720168.

Kemp, J. L., Harper, D. M. and Crosa, G. A. (1999) Use of 'functional habitats' to link ecology with morphology and hydrology in river rehabilitation, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 9 (1), pp. 159–178. DOI:10.1002/(SICI)1099-0755(199901/02)9:1<159::AID-AQC319>3.0.CO;2-M.

Kershner, J. and Snider, W. (1992) Importance of a habitat-level classification system to design instream flow studies, in: *Boon, P.J., Calow, P. & Petts, G.E. (eds.) River Conservation and management. New York, Wiley.* pp. 179–193.

Khan, J., Powell, T. and Harwood, A. (2013) Land Use in the UK, *Office for National Statistics*, pp. 19. Available from: http://www.ons.gov.uk/ons/rel/environmental/uk-environmental-accounts/2013/index.html

Khosravinia, P., Nikpour, M. R., Malekpour, A. and Hosseinzadeh Dalir, A. (2019) Effect of side slope of main channels on formation and penetration of scour hole in confluences, *River Research and Applications*, 35 (2), pp. 159–168. DOI:10.1002/rra.3398.

Kiffney, P. M., Greene, C. M., Hall, J. E. and Davies, J. R. (2006) Tributary streams create spatial discontinuities in habitat, biological productivity, and diversity in mainstem rivers, *Canadian Journal of Fisheries and Aquatic Sciences*, 63, pp. 2518–2530. DOI:10.1139/f06-138.

King, R. S., Walker, C. M., Whigham, D. F., Baird, S. J. and Back, J. A. (2012) Catchment topography and wetland geomorphology drive macroinvertebrate community structure and juvenile salmonid distributions in south-central Alaska headwater streams, *Freshwater Science*, 31 (2), pp. 341–364. DOI:10.1899/11-109.1.

Kirkby, M. (1976) Tests of the random network model, and its application to basin hydrology, *Earth Surface Processes*, 1, pp. 197–212. DOI:10.1002/esp.3290010302.

Knighton, A. D. (1980) Longitudinal changes in size and sorting of stream-bed material in four English rivers, *Geological Society of America Bulletin*, 91 (1), pp. 55–62. DOI:10.1130/0016-7606(1980)91<55:LCISAS>2.0.CO;2.

Knighton, D. (1998) *Fluvial forms and processes: a new perspective*. 3rd ed. London: Routledge.

Kohonen, T. (1982) Self-organized formation of topologically correct feature maps, *Biological Cybernetics*, 43 (1), pp. 59–69. DOI:10.1007/BF00337288.

Kohonen, T. (2001) *Self-Organizing Maps*. 3rd ed. Berlin: Springer.

Koizumi, I., Kanazawa, Y. and Tanaka, Y. (2013) The fishermen were right: experimental evidence for tributary refuge hypothesis during floods, *Zoological Science*, 30 (5), pp. 375–379. DOI:10.2108/zsj.30.375.

Kondolf, G. M., Montgomery, D. R., Piégay, H. and Schmitt, L. (2003) Geomorphic classification of rivers and streams, in: Kondolf, G. M. and Piégay, H. (eds.) *Tools in fluvial geomorphology*. Chichester: Wiley, pp. 171–204.

Kotliar, N. B. and Wiens, J. A. (1990) Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity, *Oikos*, 59 (2), pp. 253–260.

Kuemmerlen, M., Reichert, P., Siber, R. and Schuwirth, N. (2019) Ecological assessment of river networks: From reach to catchment scale, *Science of the Total Environment*, 650, pp. 1613–1627. DOI:10.1016/j.scitotenv.2018.09.019.

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z. and Kenkel, B. (2019) Classification and regression training, *Package 'caret' Version 6.0-84*. Available from: https://cran.r-project.org/package=caret

Lammert, M. and Allan, J. (1999) Assessing biotic integrity of streams: effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates, *Environmental Management*, 23 (2), pp. 257–270. DOI:10.1007/s002679900184.

Lane, E. W. (1955) The importance of fluvial morphology in hydraulic engineering, *Proceedings of American Society of Civil Engineers*, 81 (745).

Large, A. R. G. and Heritage, G. L. (2007) Terrestrial laser scanner based instream habitat quantification using a random field approach, *Proceedings of the RSPSpc Conference 2007*, pp. 4–8.

Lashermes, B. and Foufoula-Georgiou, E. (2007) Area and width functions of river networks: new results on multifractal properties, *Water Resources Research*, 43 (9), pp. 1–19. DOI:10.1029/2006WR005329.

Leal, C. G., Pompeu, P. S., Gardner, T. A., Leitão, R. P., Hughes, R. M., Kaufmann, P. R., *et al.* (2016) Multi-scale assessment of human-induced changes to Amazonian instream habitats, *Landscape Ecology*, 31 (8), pp. 1725–1745. DOI:10.1007/s10980-016-0358-x.

Leite Ribeiro, M., Blanckaert, K., Roy, A. G. and Schleiss, A. J. (2012a) Flow and sediment dynamics in channel confluences, *Journal of Geophysical Research: Earth Surface*, 117 (1). DOI:10.1029/2011JF002171.

Leite Ribeiro, M., Blanckaert, K., Roy, A. G. and Schleiss, A. J. (2012b) Hydromorphological implications of local tributary widening for river rehabilitation, *Water Resources Research*, 48 (10), pp. 1–19. DOI:10.1029/2011WR011296.

Leopold, L. B. and Maddock, T. (1953) The hydraulic geometry of stream channels and some physiographic implications, *US Government Printing Office*, 252.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S. and Bochtis, D. (2018) Machine learning in agriculture: A review, *Sensors*, 18 (8), pp. 2674. DOI:10.3390/s18082674.

Liébault, F., Gomez, B., Page, M., Marden, M., Peacock, D., Richard, D. and Trotter, C. M. (2005) Land-use change, sediment production and channel response in upland regions, *River Research and Applications*, 21 (7), pp. 739–756. DOI:10.1002/rra.880.

Lindholm, M., Grönroos, M., Hjort, J., Karjalainen, S. M., Tokola, L. and Heino, J. (2018) Different species trait groups of stream diatoms show divergent responses to spatial and environmental factors in a subarctic drainage basin, *Hydrobiologia*, 816 (1), pp. 213–230. DOI:10.1007/s10750-018-3585-0.

Lindsay, J. B. and Evans, M. G. (2008) The influence of elevation error on the morphometrics of channel networks extracted from DEMs and the implications for hydrological modelling, *Hydrological Processes*, 22, pp. 1588–1603. DOI:10.1002/hyp.6728.

Loiselle, S. A., Gasparini, D., Cunha, F., Shupe, S., Valiente, E., Baruch, A., Rocha, L., Heasley, E. and Pe, P. (2016) Micro and macroscale drivers of nutrient concentrations in urban streams in South, Central and North America, *PLoS ONE*, 11 (9), pp. 1–16. DOI:10.1371/journal.pone.0162684.

Lorenz, A. W. and Feld, C. K. (2013) Upstream river morphology and riparian land use overrule local restoration effects on ecological status assessment, *Hydrobiologia*, 704, pp. 489–501. DOI:10.1007/s10750-012-1326-3.

Mac Nally, R. (2000) Regression and model-building in conservation biology, biogeography and ecology: The distinction between – and reconciliation of – 'predictive' and 'explanatory' models, *Biodiversity and Conservation*, 9, pp. 655–671.

Macklin, M. G. and Lewin, J. (2003) River sediments, great floods and centennial-scale Holocene climate change, *Journal of Quaternary Science*, 18 (2), pp. 101–105. DOI:10.1002/jqs.751.

Maloney, K. O., Mulholland, P. J. and Feminella, J. W. (2005) Influence of catchment-scale military land use on stream physical and organic matter variables in small Southeastern Plains catchments (USA), *Environmental Management*, 35 (5), pp. 677–691. DOI:10.1007/s00267-004-4212-6.

Manfrin, A., Bombi, P., Traversetti, L., Larsen, S. and Scalici, M. (2016) A landscape-based predictive approach for running water quality assessment: A Mediterranean case study, *Journal for Nature Conservation*, 30, pp. 27–31. DOI:10.1016/j.jnc.2016.01.002.

McClain, M. E., Boyer, E. W., Dent, C. L., Gergel, S. E., Grimm, N. B., Groffman, P. M., *et al.* (2003) Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems, *Ecosystems*, 6 (4), pp. 301–312. DOI:10.1007/s10021-003-0161-9.

McGonigle, D. F., Burke, S. P., Collins, A. L., Gartner, R., Haft, M. R., Harris, R. C., *et al.* (2014) Developing Demonstration Test Catchments as a platform for transdisciplinary land management research in England and Wales., *Environmental Science: Process and Impacts*, pp. 1618–1628. DOI:10.1039/c3em00658a.

McRae, S. E., Allan, J. D., Burch, J. B., Carolina, N., Heritage, N., Arbor, A. and Arbor, A. (2004) Reach- and catchment-scale determinants of the distribution of freshwater mussels in south-eastern Michigan, U.S.A., pp. 127–142.

Messina, J. P., Brady, O. J., Golding, N., Kraemer, M. U. G., Wint, G. R. W., Ray, S. E., *et al.* (2019) The current and future global distribution and population at risk of dengue, *Nature Microbiology*. DOI:10.1038/s41564-019-0476-8.

Met Office, Hollis, D. and McCarthy, M. (2017) UKCP09: Met Office gridded and regional land surface climate observation datasets. Centre for Environmental Data Analysis.

Meybeck, M. (2003) Global analysis of river systems: from Earth system controls to Anthropocene syndromes., *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 358 (1440), pp. 1935–55. DOI:10.1098/rstb.2003.1379.

Milesi, S. V. and Melo, A. S. (2013) Conditional effects of aquatic insects of small tributaries on mainstream assemblages: position within drainage network matters, *Canadian Journal of Fisheries and Aquatic Sciences*, 71 (1), pp. 1–9.

Miller, V. C. (1953) *A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee 2*. New York: Columbia University.

Milner, V. S. and Gilvear, D. J. (2012) Characterization of hydraulic habitat and retention across different channel types; introducing a new field-based technique, *Hydrobiologia*, 694 (1), pp. 219–233. DOI:10.1007/s10750-012-1164-3.

Montgomery, D. R. (1999) Process domains and the river continuum, *Journal of the American Water Resources Association*, 35 (2), pp. 397–410.

Moore, R. V., Morris, D. G. and Flavin, R. W. (1994) Sub-set of UK digital 1:50,000 scale river centreline network., *NERC, Institute of Hydrology, Wallingford.* Available from: https://www.ceh.ac.uk/services/150000-watercourse-network

Moran, M. S., Peters, D. P. C., McClaran, M. P., Nichols, M. H. and Adams, M. B. (2008) Long-term data collection at USDA experimental sites for studies of ecohydrology, *Ecohydrology*, 1, pp. 377–398. DOI:10.1002/eco.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. and Veith, T. L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations, *Transactions of the ASABE*, 30 (3), pp. 885–900. DOI:10.1234/590.

Morris, D. G. and Flavin, R. W. (1990) A digital terrain model for hydrology, *Proceedings of the 4th International Symposium on Spatial Data Handling*, pp. 250–262.

Morris, D. G. and Flavin, R. W. (1994) Sub-set of UK 50m by 50m hydrological digital terrain model grids., *NERC, Institute of Hydrology, Wallingford.* Available from: https://www.ceh.ac.uk/services/integrated-hydrological-digital-terrain-model

Mugodo, J., Kennard, M., Liston, P., Nichols, S., Linke, S., Norris, R. H. and Lintermans, M. (2006) Local stream habitat variables predicted from catchment scale characteristics are useful for predicting fish distribution, *Hydrobiologia*, 572 (1), pp. 59–70. DOI:10.1007/s10750-006-0252-7.

Naura, M., Clark, M. J., Sear, D. A., Atkinson, P. M., Hornby, D. D., Kemp, P., *et al.* (2016) Mapping habitat indices across river networks using spatial statistical modelling of River Habitat Survey data, *Ecological Indicators*, 66, pp. 20–29. DOI:10.1016/j.ecolind.2016.01.019.

Naura, M. and Robinson, M. (1998) Principles of using River Habitat Survey to predict the distribution of aquatic species: an example applied to the native white-clawed crayfish Austropotamobius pallipes, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 8 (4), pp. 515–527.

Naura, M., Walker, J. and Maas, G. (2003) Derivation and comparison of various typologies for the WFD, *Paper Number RTT*.

Newson, M. D. (2009) *Land, water and development: sustainable and adaptive management of rivers*. 3rd ed. Abingdon: Taylor and Francis.

Newson, M. D. (2010) Understanding 'hot-spot' problems in catchments: The need for scale-sensitive measures and mechanisms to secure effective solutions for river management and conservation, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 20 (1), pp. 62–72. DOI:10.1002/aqc.1091.

Newson, M. D., Clark, M. J., Sear, D. A. and Brookes, A. (1998a) The geomorphological basis for classifying rivers, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 8 (4), pp. 415–430. DOI:10.1002/(sici)1099-0755(199807/08)8:4<415::aid-aqc276>3.0.co;2-6.

Newson, M. D., Harper, D. M., Padmore, C. L., Kemp, J. L. and Vogel, B. (1998b) A cost-effective approach for linking habitats, flow types and species requirements, *Aquatic Conservation: Marine and Freshwater Ecosystems (B)*, 8, pp. 431–446. DOI:10.1002/(SICI)1099-

0755(199807/08)8:4<431::AID-AQC302>3.0.CO;2-W.

Newson, M. D. and Newson, C. L. (2000) Geomorphology, ecology and river channel habitat: mesoscale approaches to basin-scale challenges, *Progress in Physical Geography*, 24 (2), pp. 195–217. DOI:10.1191/030913300760564724.

Office for National Statistics (2016) National Parks - Full Extent Boundaries in Great Britain (August 2016). Available from: http://geoportal.statistics.gov.uk/datasets/national-parks-august-2016-full-extent-boundaries-in-great-britain

Olden, J. D., Kennard, M. J. and Pusey, B. J. (2012) A framework for hydrologic classification with a review of methodologies and applications in ecohydrology, *Ecohydrology*, 5 (4), pp. 503–518. DOI:10.1002/eco.251.

Osborne, L. L. and Wiley, M. J. (1992) Influence of tributary spatial position on the structure of warmwater fish communities, *Canadian Journal of Fisheries and Aquatic Sciences*, 49 (4), pp. 671–681. DOI:10.1139/f92-076.

Owens, P. N., Batalla, R. J., Collins, A. J., Gomez, B., Hicks, D. M., Horowitz, A. J., *et al.* (2005) Fine-grained sediment in river systems: environmental significance and management issues, *River Research and Applications*, 21 (7), pp. 693–717. DOI:10.1002/rra.878.

Padmore, C. L. (1997) *Physical biotopes in representative channels: identification, hydraulic characterisation and application*. University of Newcastle, Thesis.

Padmore, C. L. (1998) The role of physical biotopes in determining the conservation status and flow requirements of British rivers, *Aquatic Ecosystem Health and Management*, 1 (1), pp. 25–35. DOI:10.1016/S1463-4988(98)00004-9.

Palmer, M. A. (2009) Reforming watershed restoration: Science in need of application and applications in need of science, *Estuaries and Coasts*, 32 (1), pp. 1–17. DOI:10.1007/s12237-008-9129-5.

Palmer, M. A., Menninger, H. L. and Bernhardt, E. (2010) River restoration, habitat heterogeneity and biodiversity: A failure of theory or practice?, *Freshwater Biology*, 55 (1), pp. 205–222. DOI:10.1111/j.1365-2427.2009.02372.x.

Pandolfo, T. J., Kwak, T. J., Cope, W. G., Heise, R. J., Nichols, R. B. and Pacifici, K. (2016) Species traits and catchment-scale habitat factors influence the occurrence of freshwater mussel populations and assemblages, *Freshwater Biology*, 61 (10), pp. 1671–1684. DOI:10.1111/fwb.12807.

Pardo, I. and Armitage, P. D. (1997) Species assemblages as descriptors of mesohabitats, *Hydrobiologia*, 344 (1), pp. 111–128.

Park, E. and Latrubesse, E. M. (2015) Surface water types and sediment distribution patterns at the confluence of mega rivers: The Solimões-Amazon and Negro Rivers junction, *Water Resources Research*, 51, pp. 6197–6213. DOI:10.1002/2014WR016757.

Park, Y.-S. and Hwang, S.-J. (2016) Ecological monitoring, assessment, and management in freshwater systems, *Water*, 8 (8), pp. 324. DOI:10.3390/w8080324.

Park, Y. S., Tison, J., Lek, S., Giraudel, J. L., Coste, M. and Delmas, F. (2006) Application of a self-organizing map to select representative species in multivariate analysis: A case study determining diatom distribution patterns across France, *Ecological Informatics*, 1, pp. 247–257. DOI:10.1016/j.ecoinf.2006.03.005.

Parker, C., Thorne, C. R. and Clifford, N. J. (2015) Development of ST:REAM: a reach-based stream power balance approach for predicting alluvial river channel adjustment, *Earth Surface Processes and Landforms*, 40 (3), pp. 403–413. DOI:10.1007/s13398-014-0173-7.2.

Parsons, M. and Thoms, M. C. (2007) Hierarchical patterns of physical-biological associations in river ecosystems, *Geomorphology*, 89, pp. 127–146. DOI:10.1016/j.geomorph.2006.07.016.

Paul, M. J. and Meyer, J. L. (2001) Streams in the urban landscape, *Annual Review of Ecology and Systematics*, 32, pp. 333–365.

Pearson, C. E., Ormerod, S. J., Symondson, W. O. C. and Vaughan, I. P. (2016) Resolving large-scale pressures on species and ecosystems: propensity modelling identifies agricultural effects on streams, *Journal of Applied Ecology*, 53 (2), pp. 408–417. DOI:10.1111/1365-2664.12586.

Perry, G. L. W., Wilmshurst, J. M., Mcglone, M. S. and Napier, A. (2012) Reconstructing spatial vulnerability to forest loss by fire in pre-historic New Zealand, *Global Ecology and Biogeography*, 21 (10), pp. 1029–1041. DOI:10.1111/j.1466-8238.2011.00745.x.

Perry, J. A. and Schaeffer, D. J. (1987) The longitudinal distribution of riverine benthos: a river discontinuum?, *Hydrobiologia*, 148 (3), pp. 257–268.

Perry, M. and Hollis, D. (2005) The generation of monthly gridded datasets for a range of climatic variables over the UK, *International Journal of Climatology*, 25, pp. 1041–1054. DOI:10.1002/joc.1161.

Peters, D. P. C., Burruss, N. D., Rodriguez, L. L., McVey, D. S., Elias, E. H., Pelzel-Mccluskey, A. M., *et al.* (2018) An integrated view of complex landscapes: A big data-model integration approach to transdisciplinary science, *BioScience*, 68 (9), pp. 653–669. DOI:10.1093/biosci/biy069.

Peterson, E. E., Ver Hoef, J. M., Isaak, D. J., Falke, J. A., Fortin, M. J., Jordan, C. E., *et al.* (2013) Modelling dendritic ecological networks in space: an integrated network perspective, *Ecology Letters*, 16 (5), pp. 707–719. DOI:10.1111/ele.12084.

Petts, G. E. and Amoros, C. (1996) *The Fluvial Hydrosystems*. Dordrecht: Springer.

Petts, G. E., Bickerton, M. A., Crawford, C., Lerner, D. N. and Evans, D. (1999) Flow management to sustain groundwater-dominated stream ecosystems, *Hydrological Processes*, 13, pp. 497–513. DOI:10.1002/(SICI)1099-1085(19990228)13:3<497::AID-HYP753>3.0.CO;2-S.

Piegay, H., Thevenet, A., Kondolf, G. M. and Landon, N. (2000) Physical and human factors influencing potential fish habitat distribution along a mountain river, France, *Geografiska Annaler. Series A, Physical Geography*, 82 (1), pp. 121–136. DOI:10.2307/521447.

Plug, U., Barlaup, B. T., Sternecker, K., Trepl, L. and Unfer, G. (2013) Restoration of spawning habitats of Brown Trout (Salmo trutta) in a regulated chalk stream, *River Research and Applications*, 29, pp. 172–182. DOI:10.1002/rra.

Poff, N. L., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E. and Stromberg, J. C. (1997) A paradigm for river conservation and restoration, *BioScience*, 47 (11), pp. 769–784. DOI:10.2307/1313099.

Pryde, J. K., Osorio, J., Wolfe, M. L., Heatwole, C., Benham, B. and Cardenas, A. (2007) Comparison of watershed boundaries derived from SRTM and ASTER digital elevation datasets and from a digitized topographic map, *ASABE Annual International Meeting*, 0300 (072093), pp. 1–10.

R Core Team (2018) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: https://www.r-project.org/

Rankinen, K., Cano Bernal, J. E., Holmberg, M., Vuorio, K. and Granlund, K. (2019) Identifying multiple stressors that influence eutrophication in a Finnish agricultural river, *Science of the Total Environment*, 658, pp. 1278–1292. DOI:10.1016/j.scitotenv.2018.12.294.

Raven, E. K., Lane, S. N. and Bracken, L. J. (2010) Understanding sediment transfer and morphological change for managing upland gravel-bed rivers, *Progress in Physical Geography*, 34 (1), pp. 23–45. DOI:10.1177/0309133309355631.

Raven, P. J., Fox, P., Everard, M., Holmes, N. T. H. and Dawson, F. H. (1996) River Habitat Survey: a new system for classifying rivers according to their habitat quality, in: Boon, P. J. and Howell, D. L. (eds.) *Freshwater quality: defining the indefinable?* Edinburgh: The Stationery Office, pp. 215–234.

Raven, P. J., Holmes, N. T. H., Dawson, F. H. and Everard, M. (1998) Quality assessment using River Habitat Survey data, *Aquatic Conservation: Marine and Freshwater Ecosystems (B)*, 8 (4), pp. 477–499. DOI:10.1002/(SICI)1099-0755(199807/08)8:4<477::AID-AQC299>3.0.CO;2-K.

Reid, M. A. and Thoms, M. C. (2008) Surface flow types, near-bed hydraulics and the distribution of stream macroinvertebrates, *Biogeosciences Discussions*, 5, pp. 1175–1204. DOI:10.5194/bgd-5-1175-2008.

Rhoads, B. L. (1987) Changes in stream characteristics at tributary junctions, *Physical Geography*, 8 (4), pp. 346–361.

Rice, S. P. (1998) Which tributaries disrupt downstream fining along gravel-bed rivers?, *Geomorphology*, 22 (1), pp. 39–56. DOI:10.1016/S0169-555X(97)00052-4.

Rice, S. P. (2017) Tributary connectivity, confluence aggradation and network biodiversity, *Geomorphology*, 277, pp. 6–16. DOI:10.1016/j.geomorph.2016.03.027.

Rice, S. P. and Church, M. (1998) Grain size along two gravel-bed rivers: statistical variation, spatial pattern and sedimentary links, *Earth Surface Processes and Landforms*, 23 (4), pp. 345–363. DOI:10.1002/(SICI)1096-9837(199804)23:4<345::AID-ESP850>3.0.CO;2-B.

Rice, S. P., Ferguson, R. I. and Hoey, T. B. (2006) Tributary control of physical heterogeneity and biological diversity at river confluences, *Canadian Journal of Fisheries and Aquatic Sciences*, 63 (11), pp. 2553–2566. DOI:10.1139/f06-145.

Rice, S. P., Greenwood, M. T. and Joyce, C. B. (2001) Tributaries, sediment sources, and the longitudinal organisation of macroinvertebrate fauna along river systems, *Canadian Journal of Fisheries and Aquatic Sciences*, 58 (4), pp. 824–840. DOI:10.1139/cjfas-58-4-824.

Richards, C., Haro, R., Johnson, L. B. and Host, G. (1997) Catchment and reach-scale properties as indicators of macroinvertebrate species traits, *Freshwater Biology*, 37 (1), pp. 219–230. DOI:10.1046/j.1365-2427.1997.d01-540.x.

Richards, C., Johnson, L. B. and Host, G. E. (1996) Landscape-scale influences on stream habitats and biota, *Canadian Journal of Fisheries and Aquatic Sciences*, 53 (S1), pp. 295–311. DOI:10.1139/f96-006.

Richards, K. S. (1980) A note on changes in channel geometry at tributary junctions, *Water Resources Research*, 16 (1), pp. 241–244. DOI:10.1029/WR016i001p00241.

Rinaldi, M., Gurnell, A. M., del Tánago, M. G., Bussettini, M. and Hendriks, D. (2016) Classification of river morphology and hydrology to support management and restoration, *Aquatic Sciences*, 78 (1), pp. 17–33. DOI:10.1007/s00027-015-0438-z.

Rodriguez-Iturbe, I. and Valdes, J. B. (1979) The geomorphologic structure of hydrologic response, *Water Resources Research*, 15 (6), pp. 1409–1420.

Rosgen, D. L. (1985) A stream classification system, *Riparian Ecosystems and Their Management: Reconciling Conflicting Uses. First North American Riparian Conference. Rocky Mountain Forest and Range Experiment Station*, RM-120, pp. 91–95.

Roth, N. E., Allan, J. D. and Erickson, D. L. (1996) Landscape influences on stream biotic integrity assessed at multiple spatial scales, *Landscape Ecology*, 11 (3), pp. 141–156. DOI:10.1007/BF02447513.

Rothwell, J. J., Dise, N. B., Taylor, K. G., Allott, T. E. H., Scholefield, P., Davies, H. and Neal, C. (2010) A spatial and seasonal assessment of river water chemistry across North West England, *Science of the Total Environment*, 408, pp. 841–855. DOI:10.1016/j.scitotenv.2009.10.041.

Rowntree, K. M. (1996) The hydraulics of physical biotopes-terminology, inventory and calibration, *WRC Report KV84/96*, pp. 1–51.

Rowntree, K. M. and Wadeson, R. A. (1996) Translating channel morphology into hydraulic habitat: application of the hydraulic biotope concept to an assessment of discharge related habitat change, *Proceedings of the 2nd International Association for Hydraulic Research International Symposium on Hydraulics and Habitats*, pp. A281–A292.

Rowntree, K. M. and Wadeson, R. A. (1998) A geomorphological framework for the assessment of instream flow requirements, *Aquatic Ecosystem Health and Management*, 1, pp. 125–141. DOI:10.1080/14634989808656910.

Sabo, J. L., Caron, M., Doucett, R., Dibble, K. L., Ruhi, A., Marks, J. C., Hungate, B. A. and Kennedy, T. A. (2018) Pulsed flows, tributary inputs and food-web structure in a highly regulated river, *Journal of Applied Ecology*, 55 (4), pp. 1884–1895. DOI:10.1111/1365-2664.13109.

Sayer, A. (1992) *Method in social science: A realist approach*. 2nd ed. London: Routledge.

Schindfessel, L., Creëlle, S. and De Mulder, T. (2015) Flow patterns in an open channel confluence with increasingly dominant tributary inflow, *Water*, 7, pp. 4724–4751. DOI:10.3390/w7094724.

Schmera, D., Árva, D., Boda, P., Bódis, E., Bolgovics, Á., Borics, G., *et al.* (2018) Does isolation influence the relative role of environmental and dispersal-related processes in stream networks? An empirical test of the network position hypothesis using multiple taxa, *Freshwater Biology*, 63 (1), pp. 74–85. DOI:10.1111/fwb.12973.

Schumm, S. A. (1977) *The fluvial system*. New York: John Wiley and Sons.

Schumm, S. A. and Lichty, R. W. (1965) Time, space, and causality in geomorphology, *American Journal of Science*. DOI:10.2475/ajs.263.2.110.

Seager, K., Baker, L., Parsons, H., Raven, P. J. and Vaughan, I. P. (2012) The rivers and streams of England and Wales: an overview of their physical character in 2007–2008 and changes since 1995–1996., in: Boon PJ, R. P. (ed.) *River Conservation and Management*. Chichester: John Wiley and Sons, pp. 29–43.

Sear, D. A. (1994) River restoration and geomorphology, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 4 (2), pp. 169–177. DOI:10.1002/aqc.3270040207.

Sear, D. A. (1996) The sediment system and channel stability, in: Brookes, A. and Shields, F. D. (eds.) *River channel restoration: guiding priciples for sustainable projects.* Chichester, UK: Wiley, pp. 149–177.

Sear, D. A., Armitage, P. D. and Dawson, F. H. (1999) Groundwater dominated rivers, *Hydrological Processes*, 13, pp. 255–276. DOI:10.1002/(SICI)1099-1085(19990228)13:3<255::AID-HYP737>3.0.CO;2-Y.

Sear, D. A., Newson, M. D. and Thorne, C. R. (2003) *Guidebook of applied fluvial geomorphology*. Environment Agency R&D Dissemination Centre, Swindon.

Shuker, L. J., Gurnell, A. M., Wharton, G., Gurnell, D. J., England, J., Finn Leeming, B. F. and Beach, E. (2017) MoRPh: a citizen science tool for monitoring and appraising physical habitat changes in rivers, *Water and Environment Journal*, 31 (3), pp. 418–424. DOI:10.1111/wej.12259.

Simpson, E. H. (1949) Measurement of diversity, *Nature*, 163 (688). DOI:10.1038/163688a0.

Singer, M. B. (2008) Downstream patterns of bed material grain size in a large , lowland alluvial river subject to low sediment supply, *Water Resources Research*, 44, pp. 1–7. DOI:10.1029/2008WR007183.

Singh, V. P. (1997) Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrological Processes*, 11, pp. 1649–1669. DOI:10.1002/(SICI)1099-1085(19971015)11:12<1649::AID-HYP495>3.0.CO;2-1.

Skoulikidis, N. T., Karaouzas, I. and Gritzalis, K. C. (2009) Identifying key environmental variables structuring benthic fauna for establishing a biotic typology for Greek running waters, *Limnologica*, 39 (1), pp. 56–66. DOI:10.1016/j.limno.2008.01.002.

Sliva, L. and Williams, D. D. (2001) Buffer zone versus whole catchment approaches to studying land use impact on river water quality, *Water Research*, 35 (14), pp. 3462–72. DOI:10.1016/S0043-1354(01)00062-8.

Smith, B. (2015) *River Restoration and the Water Framework Directive*. King's College London, Thesis.

Smith, B., Clifford, N. J. and Mant, J. (2014) Analysis of UK river restoration using broad-scale data sets, *Water and Environment Journal*, 28 (4), pp. 490–501. DOI:10.1111/wej.12063.

Steel, E. A., Sowder, C. and Peterson, E. E. (2016) Spatial and temporal variation of water temperature regimes on the Snoqualmie River network, *Journal of the American Water Resources Association*, 52 (3), pp. 769–787. DOI:10.1111/1752-1688.12423.

Stepinski, T. F. and Stepinski, A. P. (2005) Morphology of drainage basins as an indicator of climate on early Mars, *Journal of Geophysical Research E: Planets*, 110 (12), pp. 1–10. DOI:10.1029/2005JE002448.

Strahler, A. (1957) Quantitative analysis of watershed geomorphology, *Transactions of the American Geophysical Union*, 38 (6), pp. 913–920.

Swanson, B. J. and Meyer, G. (2014) Tributary confluences and discontinuities in channel form and sediment texture: Rio Chama, NM, *Earth Surface Processes and Landforms*, 39 (14), pp. 1927–1943. DOI:10.1002/esp.3586.

Tadaki, M., Brierley, G. and Cullum, C. (2014) River classification: theory, practice, politics, *Wiley Interdisciplinary Reviews: Water*, 1 (4), pp. 349–367. DOI:10.1002/wat2.1026.

Taigel, S. H. (2016) *Utilising spatial technologies to support the Catchment Based Approach to landscape management*. University of East Anglia.

Tetzlaff, D., Soulsby, C., Bacon, P. J., Youngson, A. F., Gibbins, C. and Malcolm, I. A. (2007) Connectivity between landscapes and riverscapes - a unifying theme in integrating hydrology and ecology in catchment science?, *Hydrological Processes*, 21, pp. 1385–1389. DOI:10.1002/hyp.

Therneau, T. M. and Atkinson, E. J. (1997) An introduction to recursive partitioning using the RPART routines, *Mayo Clinic*. Available from: https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf [Accessed 1 July 2019].

Thompson, C., Croke, J. and Takken, I. (2008) A catchment-scale model of mountain stream channel morphologies in southeast Australia, *Geomorphology*, 95 (3–4), pp. 119–144. DOI:10.1016/j.geomorph.2007.05.015.

Townsend, C. R. (1989) The patch dynamics concept of stream community ecology, *Journal of the North American Benthological Society*, 8 (1), pp. 36–50.

Townsend, C. R., Doledec, S., Norris, R., Peacock, K. and Arbuckle, C. (2003) The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction, *Freshwater Biology*, 48, pp. 768–785.

Trimble, S. W. and Mendel, A. C. (1995) The cow as a geomorphic agent - a critical review, *Geomorphology*, 13 (1–4), pp. 233–253. DOI:10.1016/0169-555X(95)00028-4.

Turner, M. G. and Ruscher, C. L. (1988) Changes in landscape patterns in Georgia, USA, *Landscape Ecology*, 1 (4), pp. 241–251. DOI:10.1007/BF00157696.

Udvardy, M. F. D. (1959) Notes on the ecological concepts of habitat, biotope and niche, *Ecology*, 40 (4), pp. 725–728.

UKTAG (2003) Guidance on Typology for Rivers for Scotland, England and Wales, *UK Technical Advisory Group on the Water Framework Directive: Work Program - Task 2a. Typology for Rivers*, pp. 1–4.

Vander Vorste, R., McElmurray, P., Bell, S., Eliason, K. M. and Brown, B. L. (2017) Does stream size really explain biodiversity patterns in lotic systems? A call for mechanistic explanations, *Diversity*, 9 (3), pp. 1–21. DOI:10.3390/d9030026.

Vannote, R., Minshall, G., Cummins, K., Sedell, J. and Cushing, C. (1980) The River Continuum Concept, *Canadian Journal of Fisheries and Aquatic Sciences*, 37 (1), pp. 130–137. DOI:10.1139/f80-017.

Vaughan, I. P., Diamond, M., Gurnell, A. M., Hall, K. A., Jenkins, A., Milner, N. J., *et al.* (2009) Integrating ecology with hydromorphology: a priority for river science and management, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 19 (1), pp. 113–125. DOI:10.1002/aqc.895.

Vaughan, I. P., Merrix-Jones, F. L. and Constantine, J. A. (2013) Successful predictions of river characteristics across England and Wales based on ordination, *Geomorphology*, 194, pp. 121–131. DOI:10.1016/j.geomorph.2013.03.036.

Vaughan, I. P., Noble, D. G. and Ormerod, S. J. (2007) Combining surveys of river habitats and river birds to appraise riverine hydromorphology, *Freshwater Biology*, 52 (11), pp. 2270–2284. DOI:10.1111/j.1365-2427.2007.01837.x.

Vaughan, I. P. and Ormerod, S. J. (2010) Linking ecological and hydromorphological data: approaches, challenges and future prospects for riverine science, *Aquatic Conservation: Marine and Freshwater Ecosystems*, 20 (1), pp. 125–130.

Ver Hoef, J. M. and Peterson, E. E. (2010) A moving average approach for spatial statistical models of stream networks, *Journal of the American Statistical Association*, 105 (489), pp. 6–18. DOI:10.1198/jasa.2009.ap08248.

Vesanto, J. (2000) Neural network tool for data mining: SOM toolbox, in: *Proceedings of symposium on tool environments and development methods for intelligent systems*. pp. 184–196.

Wadeson, R. A. and Rowntree, K. M. (1998) Application of the hydraulic biotope concept to the classification of instream habitats, *Aquatic Ecosystem Health and Management Management*, 1 (2), pp. 143–157. DOI:10.1080/14634989808656911.

Waite, I. R., Munn, M. D., Moran, P. W., Konrad, C. P., Nowell, L. H., Meador, M. R., Van Metre, P. C. and Carlisle, D. M. (2019) Effects of urban multi-stressors on three stream biotic assemblages, *Science of the Total Environment*, 660, pp. 1472–1485. DOI:10.1016/j.scitotenv.2018.12.240.

Walley, W. J., Martin, R. W. and O'Connor, M. A. (1999) Self-Organising Maps for the classification and diagnosis of river quality from biological and environmental data, *International Symposium on Environmental Software Systems*, pp. 27–41. DOI:10.1007/978-0-387-35503-0.

Walsh, C. J., Roy, A. H., Feminella, J. W., Cottingham, P. D., Groffman, P. M. and Raymond, P. M. (2005) The urban stream syndrome: current knowledge and the search for a cure, *Journal of the North American Benthological Society*, 24 (3), pp. 706–723. DOI:10.1016/B978-0-12-801231-4.00027-6.

Wang, L. Z., Lyons, J., Kanehl, P. and Bannerman, R. (2001) Impacts on stream habitat and fish across multiple spatial scales, *Environmental Management*, 28 (2), pp. 255–266. DOI:10.1007/s002670010222.

Wehrens, R. and Kruisselbrink, J. (2018) Supervised and unsupervised Self-Organising Maps, *Package 'kohonen' Version 3.0.7*.

Weiss, A. D. (2001) Topographic position and landforms analysis, *Poster Presentation. ESRI User Conference, San Diego, CA*.

Werritty, A. (1992) Downstream fining in a gravel bed river in Southern Poland: lithological controls and the role of abrasion, in: *Dynamics of gravel bed rivers*. Chichester, UK: Wiley, pp. 333–350.

Wessels, K. J., Van Jaarsveld, A. S., Grimbeek, J. D. and Van Der Linde, M. J. (1998) An evaluation of the gradsect biological survey method, *Biodiversity and Conservation*, 7 (8), pp. 1093–1121. DOI:10.1023/A:1008899802456.

Wharton, G. (1994) Progress in the use of drainage network indices for rainfall-runoff modelling and runoff prediction, *Progress in Physical Geography*, 18 (4), pp. 539–557.

Wharton, G., Mohajeri, S. H. and Righetti, M. (2017) The pernicious problem of streambed colmation: a multi-disciplinary reflection on the mechanisms, causes, impacts, and management challenges, *WIREs Water*, e1231, pp. 1–17. DOI:10.1002/wat2.1231.

White, M. S., Tavernia, B. G., Shafroth, P. B., Chapman, T. B. and Sanderson, J. S. (2018) Vegetative and geomorphic complexity at tributary junctions on the Colorado and Dolores Rivers: a blueprint for riparian restoration, *Landscape Ecology*, 33 (12), pp. 2205–2220. DOI:10.1007/s10980-018-0734-9.

Wiens, J. A. (2002) Riverine landscapes: taking landscape ecology into the water, *Freshwater Biology*, 47, pp. 501–515.

Willgoose, C. and Hancock, G. (1998) Revisiting the hypsometric curve as an indicator of form and process in transport-limited catchment, *Earth Surface Processes and Landforms*, 23, pp. 611–623. DOI:10.1002/(SICI)1096-9837(199807)23:7<611::AID-ESP872>3.0.CO;2-Y.

Wolman, M. G. and Miller, J. P. (1960) Magnitude and Frequency of Forces in Geomorphic Processes, *The Journal of Geology*, 68 (1), pp. 54–74. DOI:10.1086/626637.

Woodcock, T., Mihuc, T., Romanowicz, E. and Allen, E. (2006) Land-use effects on catchment- and patch-scale habitat and macroinvertebrate responses in the adirondack uplands, *American Fisheries Society Symposium*, 2006 (48), pp. 395–411.

Woodward, G. and Hildrew, A. (2002) Food web structure in riverine landscapes, *Freshwater Biology*, 47 (4), pp. 777–798.

Wright, J. F., Moss, D., Armitage, P. D. and Furse, M. T. (1984) A preliminary classification of running-water sites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data, *Freshwater Biology*, 14 (3), pp. 221–256.

DOI:10.1111/j.1365-2427.1984.tb00039.x.

Zavadil, E., Stewardson, M. J., Turner, M. E. and Ladson, A. (2012) An evaluation of surface flow types as a rapid measure of channel morphology for the geomorphic component of river condition assessment, *Geomorphology*, 139–140, pp. 303–312. DOI:10.1016/j.geomorph.2011.10.034.

Zhou, T., Wu, J. and Peng, S. (2012) Assessing the effects of landscape pattern on river water quality at multiple scales: a case study of the Dongjiang River watershed, China, *Ecological Indicators*, 23, pp. 166–175. DOI:10.1016/j.ecolind.2012.03.013.

Zorn, T. G. and Wiley, M. J. (2006) Influence of landscape characteristics on local habitat and fish biomass in streams of Michigan's Lower Peninsula, *American Fisheries Society Symposium*, 2006 (48), pp. 375–393.

Appendices

Appendix numbers reflect the chapter in which the Appendix is first mentioned.

Appendix 2A. Papers selected from the quick scoping review.

	Author	Title	Location
1	(Lindholm *et al.*, 2018)	Different species trait groups of stream diatoms show divergent responses to spatial and environmental factors in a subarctic drainage basin	Finland
2	(de Castro *et al.*, 2017)	Landscape variables influence taxonomic and trait composition of insect assemblages in Neotropical savanna streams	Brazil
3	(Benone *et al.*, 2017)	Regional controls on physical habitat structure of amazon streams	Brazil
4	(Pandolfo *et al.*, 2016)	Species traits and catchment-scale habitat factors influence the occurrence of freshwater mussel populations and assemblages	NC, USA
5	(Leal *et al.*, 2016)	Multi-scale assessment of human-induced changes to Amazonian instream habitats	Brazil
6	(Manfrin *et al.*, 2016)	A landscape-based predictive approach for running water quality assessment: A Mediterranean case study	Italy
7	(Pearson *et al.*, 2016)	Resolving large-scale pressures on species and ecosystems: propensity modelling identifies agricultural effects on streams	UK
8	(Jähnig *et al.*, 2015)	Community-environment relationships of riverine invertebrate communities in central Chinese streams	China
9	(Buendia *et al.*, 2013)	Reach and catchment-scale influences on invertebrate assemblages in a river with naturally high fine sediment loads	Spain
10	(de Paula *et al.*, 2013)	Influence of forest cover on in-stream large wood in an agricultural landscape of south eastern Brazil: a multi-scale analysis	Brazil
11	(Atkinson *et al.*, 2012)	Scale-dependent longitudinal patterns in mussel communities	AR and OK, USA
12	(King *et al.*, 2012)	Catchment topography and wetland geomorphology drive macroinvertebrate community structure and juvenile salmonid distributions in south-central Alaska headwater streams	AK, USA
13	(Hutchens *et al.*, 2009)	Multi-scale mechanistic indicators of Midwestern USA stream macroinvertebrates	MI and MN, USA
14	(Skoulikidis *et al.*, 2009)	Identifying key environmental variables structuring benthic fauna for establishing a biotic typology for Greek running waters	Greece
15	(Thompson *et al.*, 2008)	A catchment-scale model of mountain stream channel morphologies in southeast Australia	Australia
16	(Frappier and Eckert, 2007)	A new index of habitat alteration and a comparison of approaches to predict stream habitat conditions	NH, USA
17	(Parsons and Thoms, 2007)	Hierarchical patterns of physical-biological associations in river ecosystems	Australia
18	(Karaouzas *et al.*, 2007)	Land use effects on macroinvertebrate assemblages and stream quality along an agricultural river basin	Greece
19	(Mugodo *et al.*, 2006)	Local stream habitat variables predicted from catchment scale characteristics are useful for predicting fish distribution	Australia
20	(Zorn and Wiley, 2006)	Influence of landscape characteristics on local habitat and fish biomass in streams of Michigan's Lower Peninsula	MI, USA
21	(Woodcock *et al.*, 2006)	Land-use effects on catchment- and patch-scale habitat and macroinvertebrate responses in the Adirondack uplands	NY, USA
22	(Maloney *et al.*, 2005)	Influence of catchment-scale military land use on stream physical and organic matter variables in small south eastern plains catchments (USA)	GA, USA
23	(Emery *et al.*, 2004)	Characteristics and controls of gravel-bed riffles: An analysis of data from the river-habitat survey	UK
24	(Death and Joy, 2004)	Invertebrate community structure in streams of the Manawatu-Wanganui region, New Zealand: the roles of catchment versus reach scale influences	New Zealand
25	(Feld, 2004)	Identification and measure of hydromorphological degradation in Central European lowland streams	Europe
26	(McRae *et al.*, 2004)	Reach- and catchment-scale determinants of the distribution of freshwater mussels (*Bivalvia: Unionidae*) in south-eastern Michigan, USA	MI, USA
27	(Townsend *et al.*, 2003)	The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction	New Zealand
28	(Davies *et al.*, 2000)	Prediction and assessment of local stream habitat features using large-scale catchment characteristics	Australia
29	(Richards *et al.*, 1997)	Catchment and reach-scale properties as indicators of macroinvertebrate species traits	MI, USA

Appendix 3A. Selecting the number of SOM clusters

Hierarchical clustering was applied to the SOM output to identify typology classes. The Davies-Bouldin index, a measure of clustering quality, indicates that 5, 7 or 15 clusters are preferable as a result of the low index values (**Figure 3A1a**). The index suggests five clusters are statistically optimal, but this number was not selected as the complexity of catchment characteristics that influence river functioning (**Table 3.2**) is not sufficiently captured for management purposes. For example, if five clusters are selected, groundwater dominated waterbodies and highly seasonal catchments would not be classified into separate waterbody types (**Figure 3Ab**). On the other hand, fifteen clusters reflect subtle variations within types (as indicated by high U-matrix values; **Figure 3.1b**) producing a finer classification, primarily along the vertical gradient of the grid (**Figure 3Ab**). This additional level of detail does not add much further representation of catchment controls useful for management and so was considered too complicated. Therefore, seven clusters are selected to create seven waterbody types (**Figure 3.1c**).

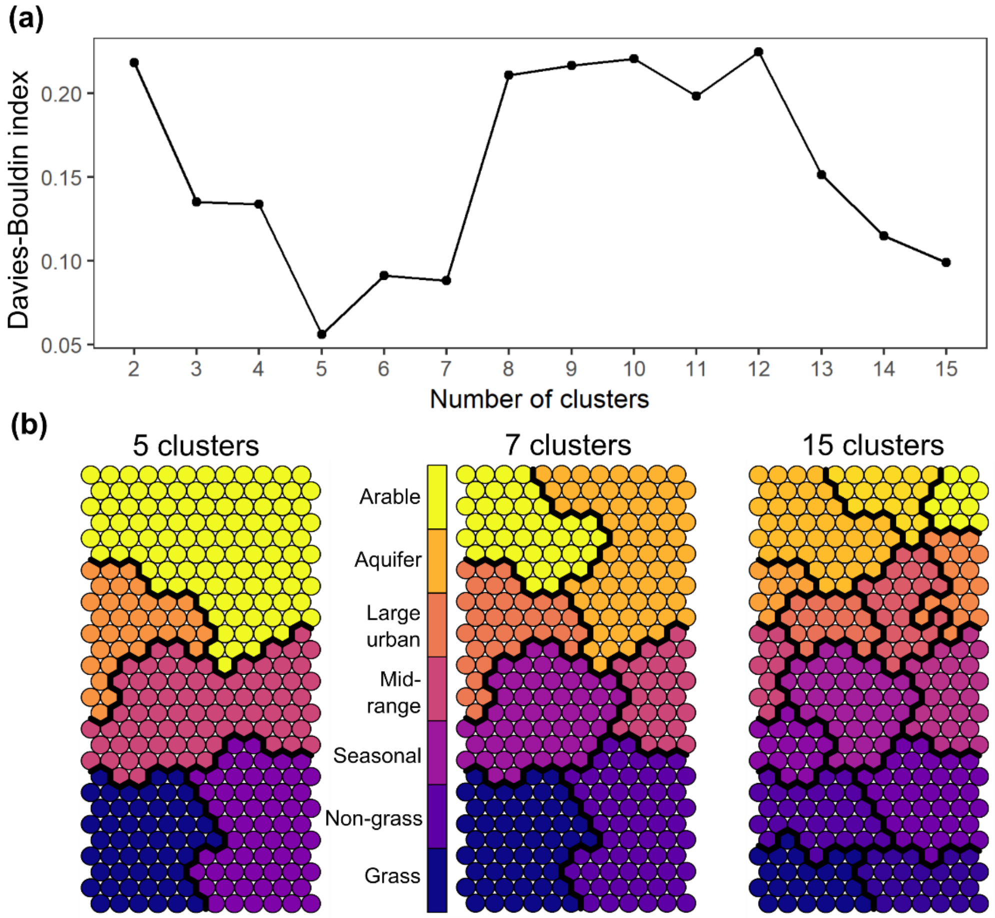


Figure 3A. *Identifying the appropriate number of clusters to represent waterbody types: (a) Low Davies-Bouldin Index values indicate the optimum number of clusters. (b) Boundaries of 5, 7 and 15 waterbody types, the numbers of clusters with the lowest Davies-Bouldin index values, plotted on the SOM grid from* ***Figure 3.1****. Seven types were selected based on expert judgement for the intended purpose, described in the text. Names of the selected seven waterbody types reflect the characteristics of the type, see* ***Figure 3.1****.*

Appendix 3B. Impact of grid-shape on SOM outputs

SOM maps were produced for different grid dimensions (**Figure 3B**) to identify how much of an impact this has on the results. All grids have 336 cells. The number of cells was estimated in the same manner as the paper where the number of cells is approximately equal to $5\sqrt{N}$ where N is the number of samples (Vesanto, 2000). The grid with 12 x 28 dimensions reflect the ratio of the two largest eigenvalues of the input variables, the method of grid dimension estimation used in the paper (Park *et al.*, 2006). These grid dimensions produce maps where all cells contain waterbodies, a clear gradient from low to high elevation (from top to bottom of the map) and a grouping of hard rock geology.

In the squarer grid (16 x 21), not all cells are populated by waterbodies (indicated by two grey cells). The elevation heatmap also shows a gradient from low to high elevation (from top right to bottom left), however there are anomalous high-elevation cells in the top right corner. This is also the case for many of the continuous variables not shown in **Figure 3B**. The hard rock heatmap is also split into three distinct categories at the corners of the map, suggesting again that there may be anomalies in grouping the most similar waterbodies together.

The elongated rectangular grid (8 x 42) is the worst SOM grid shape. The count map shows that most waterbodies are grouped within five cells containing up to 150 waterbodies so the map cannot represent variation in waterbody characteristics. The elevation and hard rock heatmaps are scattered with no clear patterns. It is therefore concluded that the original 12 x 28 grid is the best representation of the catchment characteristics.

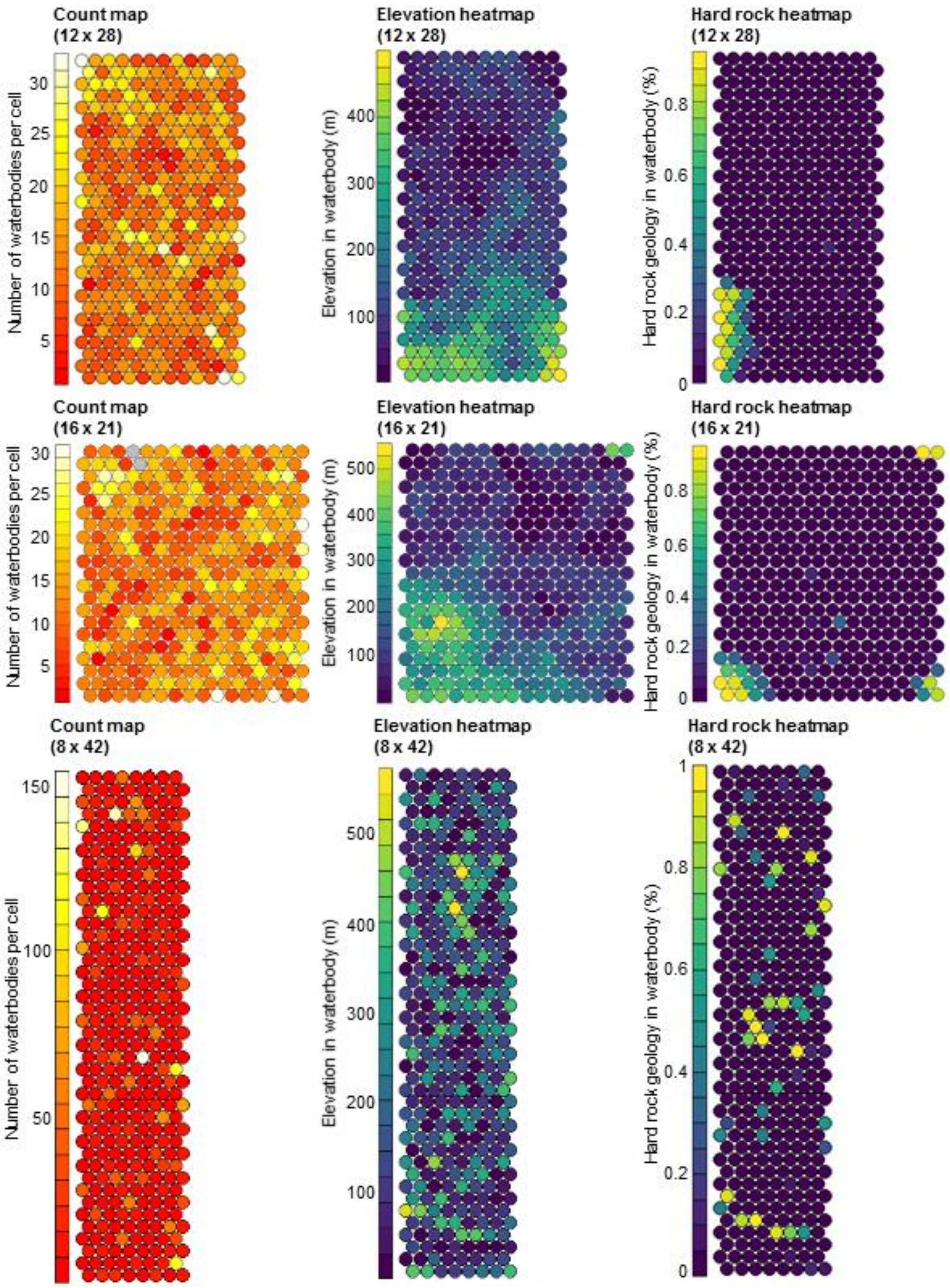


Figure 3B. *An example of SOM results for different sizes of lattice grid: 12 x 28, 16 x 21 and 8 x 42. Results are count maps (left column) which show the number of waterbodies classified in each output neuron and heatmaps showing the values of a continuous input variable (elevation; centre column) and a categorical input variable (hard rock geology; right column) for each output neuron. Only two variable heatmaps are displayed for clarity showing how each SOM map deals with continuous and categorical variables*

Appendix 3C. Model code for SOM clusters

The R code below can be used to conduct SOM analysis on the 22 catchment characteristics from the dataset 'charac' and conduct hierarchical clustering.

```
#data
## One row for each waterbody, first column waterbody ID,
## remaining columns each of the 22 catchment characteristics
head(charac)

#Preparation for SOM
data=charac[,-c(1)]
n_iterations = 10000
recalculate_map = T
recalculate_no_clusters = T

data_list=list()
distances = vector()
data.nam<-names(data)
data_list[['data.nam']] = scale(data[,data.nam])
distances = c( distances, 'euclidean')

#Create grid to map SOM
som_grid = kohonen::somgrid(xdim=12,ydim=28,topo="hexagonal")

#Conduct SOM
som_model = kohonen::supersom(data_list,
                              grid=som_grid,
                              rlen= n_iterations,
                              alpha = 0.05,
                              normalizeDataLayers = FALSE,
                              dist.fcts = distances)

#Hierarchical clustering
codes = tibble(layers = names(som_model$codes),codes = som_model$codes ) %>%
  mutate(codes = purrr::map(codes, as_tibble)) %>%
  spread(key = layers, value = codes) %>%
  apply(1, bind_cols) %>% .[[1]] %>% as_tibble()
dist_m<-dist(codes) %>% as.matrix()
dist_on_map<-kohonen::unit.distances(som_grid)
dist_adj = dist_m ^ dist_on_map #produce distance matrix for clustering

clust_adj = hclust(as.dist(dist_adj), 'ward.D2')
som_cluster_adj = cutree(clust_adj, 7)

#Un-scale variables for plotting heatmaps
x<-1:22

for(i in x){
  df<-aggregate(as.numeric(data[,i]), by=list(som_model$unit.classif),
       FUN=mean, simplify=TRUE)[,2]
  assign(paste("unscaled",i, sep=""),df)
}

dfs<-list(unscaled3,unscaled4,unscaled6,unscaled5,unscaled7,unscaled8,
       unscaled2, unscaled1, unscaled13, unscaled10, unscaled11, unscaled9,
       unscaled12, unscaled14, unscaled15, unscaled16, unscaled17, unscaled21,
        unscaled18, unscaled22, unscaled19, unscaled20)
nums<-c(3,4,6,5,7,8,2,1,13,10,11,9,12,14,15,16,17,21,18,22,19,20)

#plot heatmaps
par(mfrow=c(5,5),cex=0.9, mex=1)
plot(som_model, type="counts", palette.name=plasma, heatkeywidth=1,
       main="Count")
plot(som_model, type="dist.neighbours", palette.name=plasma, heatkeywidth=1,
       main="U-matrix")
for(i in dfs){
  for(j in nums){
    plot(som_model, type="property", property=i, main=names(data)[j],
palette.name=viridis,heatkeywidth = 1)
  }
}
plot(som_model, type = "property", property=som_cluster_adj, palette.name=
plasma, pchs = NA, main="Clusters",heatkeywidth = 1)
add.cluster.boundaries(som_model, som_cluster_adj)
dev.off()
```

Appendix 4A. Method for removal of anabranches

The paper uses CEH's 1:50,000 blue-line river network map (Moore *et al.*, 1994) for calculating the network density metrics. The paper describes the removal of anabranches from the network to obtain a dendritic network required for creation of the network density metrics (Section 4.2.2.2). Here, this process is described in more detail.

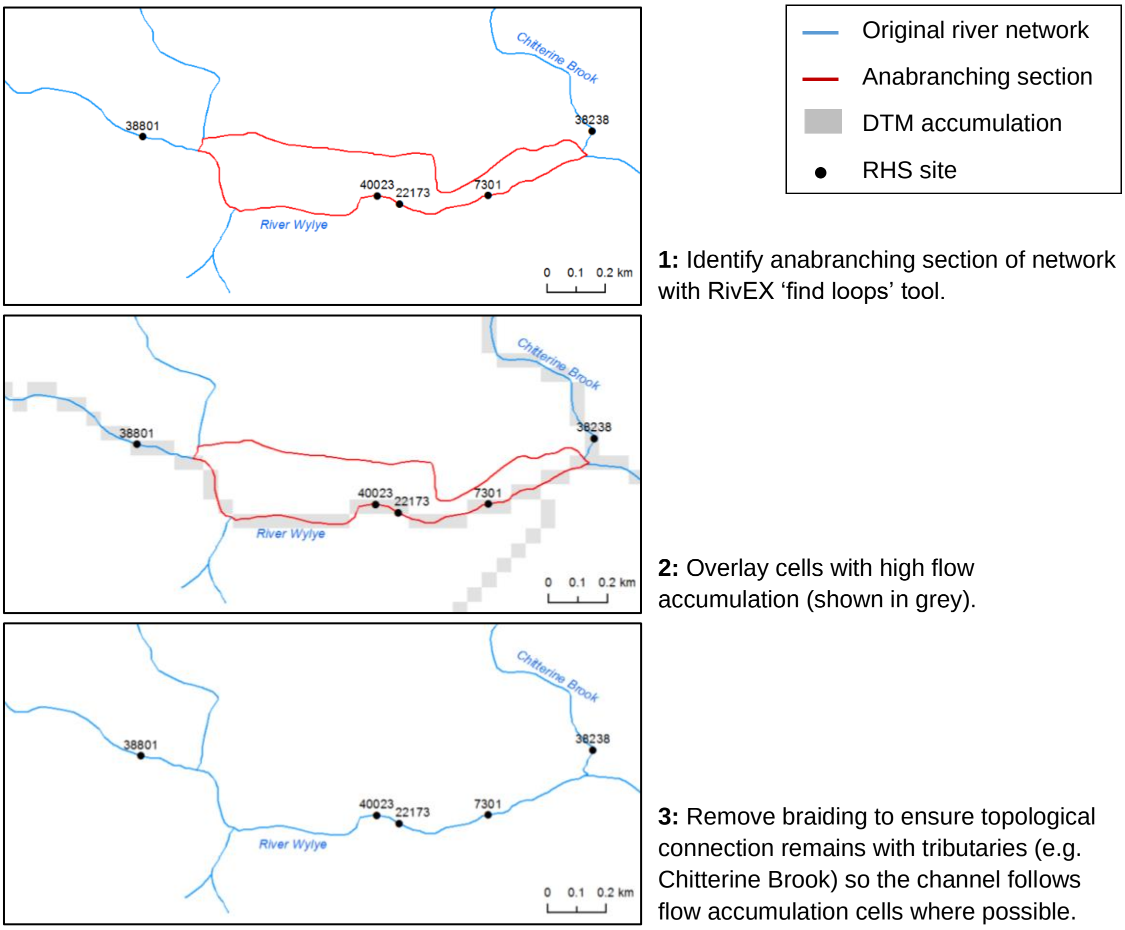


Figure 4A. *Process of manually removing braiding from section of River Wylye, a tributary of the Avon in Hampshire, including RHS surveys as black points.*

The topology of the network was first quality controlled using the RivEX (Hornby, 2010) extension for ArcMap, by ensuring there is no double digitation and self-intersection of network links and no sources or cycles within the network. Anabranching channels are identified and corrected using the steps in **Figure 4A**. The topology of the network was maintained by ensuring that tributaries still connected to the main channel. Links in the network that overlay cells with high flow accumulation (identified from the Integrated Hydrological DTM; Morris and Flavin, 1994) are the preserved as they are topographically distinct so likely to be the main channel. The location of RHS sites on the network is also considered and channels that intersect sites are retained where possible.

Appendix 5A. Anabranch removal protocol for river network in England

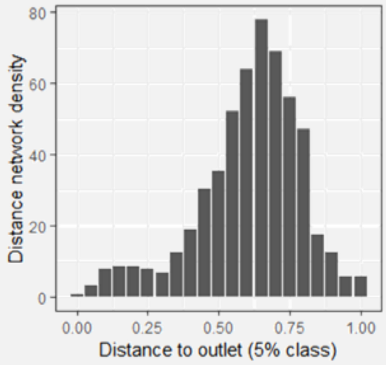
When the network metrics were developed in Chapter 4, anabranches in the networks of the DTCs were manually removed (**Appendix 4A**). However, this is impractical when dealing with the entire river network of England. Therefore, RivEX (Hornby, 2010) is used in ArcGIS v10.3 to identify anabranches in the network using the 'find loops' function. Anabranches made up of over 50 links are manually simplified to a dendritic network structure based on flow pathways identified from the IHDTM (Morris and Flavin, 1990, 1994) (as per the manual method). The number of links in anabranches with under 50 links was set to one to reduce the number of excess links influencing network density calculations but retaining the presence of the network in these areas.

Appendix 5B. Code for calculating network density metrics for catchments in England

The R function below can be used to calculate network density for catchments in England from the dataset 'net.data' where each row contains the information for an individual link extracted from RivEX.

```r
#data
head(net.data)
##CatchID - individual ID for each catchment
##RivID - individual ID for each link in the river network
##LoopID - individual ID for each loop identified with the RivEX'LoopID' tool
##ElevM - elevation at the downstream end of each link in m
##DistKM - distance from the downstream end of each link to the mouth in km

#network density function
netdens.func<-function(data,var){

  names(data)[names(data)==var]<- "var"##rename the selected variable

  ##calculate 5% distance or elevation bands
  max<-aggregate(data$var, list(data$CatchID), max)
  min<-aggregate(data$var, list(data$CatchID), min)
  names(max)<-c("CatchID", "max" )
  names(min)<-c("CatchID", "min" )
  join<-left_join(data,max,by="CatchID") %>% left_join(.,min,by="CatchID")
  join$width<-0.05*(join$max-join$min)
  join$norm<-(join$var-join$min)/(join$max-join$min)
  join$class<-round_any(join$norm, 0.05, ceiling)

  ##Produce number of links in each class for dendritic network by calculating
  ##the number of links in each band that are not dendritic (loops)
  ##while retaining a single link in each 'loop' to retaining network topology
  noloops<-join[(join$LoopID==0),]
  noloops<-noloops %>% group_by(CatchID,class) %>%
    summarise(numnolooplinks = n(), width=max(width))

  onlyloops<-join[!(join$LoopID==0),]
  onlyloops<-onlyloops %>% group_by(CatchID,class) %>%
    summarise(numonlyloops=n_distinct(LoopID))

  net<-full_join(noloops, onlyloops, by=c("CatchID","class"))
  net$numonlyloops[is.na(net$numonlyloops)]<-0
  net$numlinks<-net$numnolooplinks+net$numonlyloops

  ##Calculate network density
  net$netdensity<-(net$numlinks/net$width)
  net<-net %>% group_by(CatchID) %>% mutate(normnetdensity=
      (netdensity-min(netdensity))/(max(netdensity)-min(netdensity)))
  net<-net[,c(1,2,7,8)]
  return(net)}
```

```
#Calculate distance and elevation network density
distdens<-netdens.func(data=net.data, var="DistKM")
elevdens<-netdens.func(data=net.data, var="ElevM")

#Select only one catchment to plot density histogram,
## e.g. Tamar
eg.distdens<-netdens.func(data=tamar, var="DistKM")

ggplot(eg.distdens, aes(x=class,y=netdensity))+
  geom_bar(stat="identity")+
  xlab("Distance to outlet (5% class)")+
  ylab("Distance network density")+
  theme_bw()
```



Appendix 5C. Tributary property extraction using 100m buffer

Buffer widths from 100m to several hundred meters on both banks are common spatial scales used to relate landscape variables to stream condition (Allan, 2004). A buffer width of 100m was chosen to minimise buffer overlap as the network converges (**Figure 5Ci**). While catchment controls within the buffer do not capture the entire catchment effect, it is a consistent method that is more practical than computing a watershed boundary for every link in England's river network, a process that is dependent on the accuracy of the DEM (Lindsay and Evans, 2008; Pryde *et al.*, 2007) and would require rigorous and time consuming quality control at this broad level.

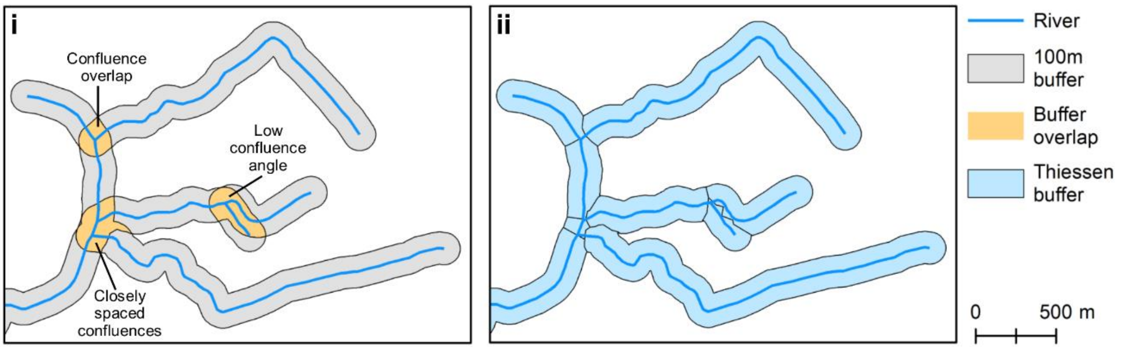


Figure 5C. *Example of the 100m buffer around the stream network used to extract upstream properties of each tributary: (i) areas of overlap in the buffer indicated; (ii) Thiessen polygons used to remove the overlap.*

While the choice of a 100m buffer minimised buffer overlap, overlap is still present at confluences and is worse where tributaries are close together or if the incoming tributaries have a low confluence angle (**Figure 5Ci**). Therefore, Thiessen polygons are calculated to segment the buffer with the area nearest in Euclidean distance to each link in the network (**Figure 5Cii**). For each tributary entering the confluence, the total area covered by each land cover or geology class within the Thiessen segments of the 100m buffer is summed so percentage cover can be compared between the incoming tributaries.

Appendix 6A. Different pruning methods for single regression trees

Different pruning methods have negligible effect on predictive capacity of single tree models. However, pruning method 1 (tree pruned at number of splits at minimum cross-validation error) provides more accurate predictions according to the model evaluators (Method 1 R^2 = 0.08-0.41 compared to Method 2 R^2 = 0.07-0.39) and of extreme observed values than the more conservative pruning method 2 (tree pruned at lowest number of splits within 1 standard deviation of minimum cross-validation error) (**Figure 6A**). Therefore, single tree models plotted in **Figure 6.6** are based on pruning method 1. However, method 1 may overfit the model to the training data so the branches retained in pruning method 2, and therefore have more confidence in, are indicated in **Figure 6.6**.

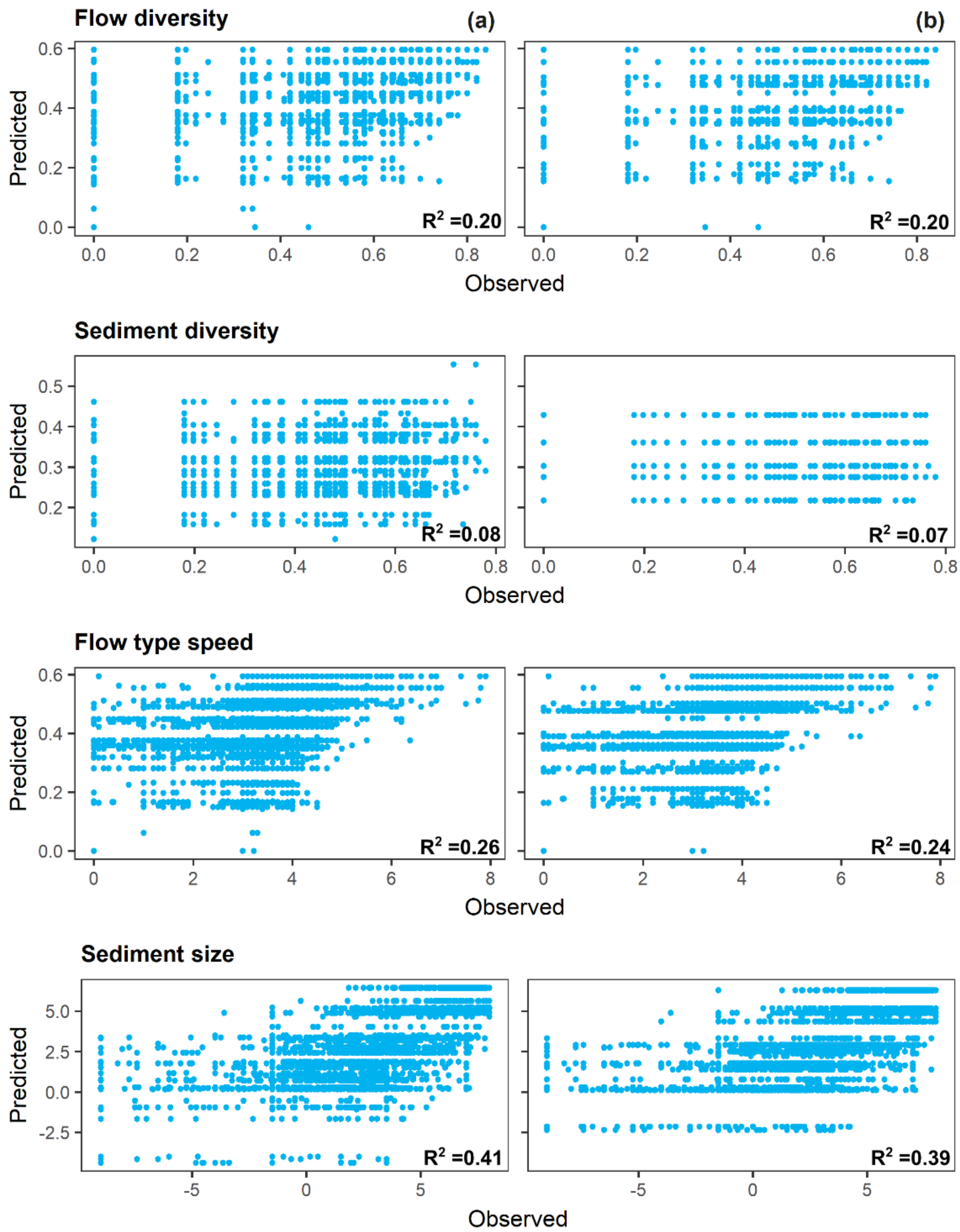


Figure 6A. *Observed vs predicted values for each habitat index for (a) pruning method 1 and (b) pruning method 2. R^2 values presented in corner of each plot.*

Appendix 6B. Model tuning parameters

Varying model tuning parameters (LR – learning rate and TC – tree complexity) to identify optimum number of trees for building the final BRT model (**Table 6B**). Optimal models must have at least 1000 trees according to Elith *et al.*'s (2008) rule of thumb.

Table 6B. *Number of trees and deviance of BRT models built with different combinations of model tuning parameters (LR and TC). The optimal models, with the lowest mean deviance, are highlighted in yellow.*

(a) Flow diversity

LR	*TC*	Number of trees	Mean deviance	Std. dev. deviance
0.005	*3*	5450	0.0408	0.0003
0.01	*3*	4550	0.0406	0.0006
0.05	*3*	1750	0.0404	0.0004
0.005	*5*	4950	0.0405	0.0004
0.01	*5*	3800	0.0403	0.0004
0.05	*5*	1200	0.0403	0.0004
0.005	*7*	4300	0.0404	0.0004
0.01	*7*	3400	0.0401	0.0005
0.05	*7*	750	0.0403	0.0006

(b) Sediment diversity

LR	*TC*	Number of trees	Mean deviance	Std. dev. deviance
0.005	*3*	3850	0.0502	0.0003
0.01	*3*	3150	0.0501	0.0003
0.05	*3*	1050	0.0500	0.0003
0.005	*5*	3800	0.0500	0.0003
0.01	*5*	2950	0.0499	0.0004
0.05	*5*	950	0.0499	0.0004
0.005	*7*	3700	0.0499	0.0004
0.01	*7*	2600	0.0498	0.0003
0.05	*7*	700	0.0498	0.0004

(c) Flow type speed

LR	*TC*	Number of trees	Mean deviance	Std. dev. deviance
0.005	*3*	8100	0.8636	0.0116
0.01	*3*	6950	0.8551	0.0151
0.05	*3*	4100	0.8450	0.0135
0.005	*5*	8450	0.8453	0.0123
0.01	*5*	6750	0.8381	0.0160
0.05	*5*	2800	0.8362	0.0122
0.005	*7*	7150	0.8402	0.0106
0.01	*7*	5850	0.8328	0.0151
0.05	*7*	2050	0.8364	0.0146

**Optimal trees never reached*

(d) Sediment size

LR	*TC*	Number of trees	Mean deviance	Std. dev. deviance
0.005	*3*	10000*	6.3672	0.0963
0.01	*3*	7650	6.2688	0.1029
0.05	*3*	3600	6.1798	0.1056
0.005	*5*	8750	6.1804	0.1112
0.01	*5*	6950	6.0993	0.1110
0.05	*5*	2200	6.1270	0.0996
0.005	*7*	8100	6.1228	0.1043
0.01	*7*	5250	6.1325	0.1464
0.05	*7*	1850	6.0888	0.0853

Appendix 6C. Code for calculating BRT model

The R code below used to produce the BRT models. The code to identify the optimal tree settings was adapted from code in Elith *et al.* (2008) using the '*dismo*' v.1.1.4 package (Hijmans *et al.*, 2017) and the selected models were built in the '*gbm*' v.2.1.5 package (Brandon *et al.*, 2019). One example of the flow type diversity index is provided here but the method was repeated for all four habitat indices.

```
#data
## Rows for each RHS site, columns contain RHS site ID, flow diversity at the
## site and variables from reach to catchment-levels
head(data)

#Choosing optimal number of trees using 'dismo' package by Elith et al. (2008)
## learning rate and tree complexity varied to determine optimum tree number

hyper_grid <- expand.grid(
  learning.rate = c(0.005,.01, .05),
  tree.complexity = c(3, 5, 7),
  optimal_trees = 0,
  mean_dev = 0,
  std_dev = 0
)

for(i in 1:nrow(hyper_grid)) {
  set.seed(123)

  # train model
  gbm.tune <- gbm.step(
    data=dat.SS,
    gbm.x = 3:38, #predictors
    gbm.y = 2,     #response
    family = "gaussian",
    tree.complexity = hyper_grid$tree.complexity[i],
    learning.rate = hyper_grid$learning.rate[i],
    bag.fraction = 0.75,     #75% data used for training
    plot.main = F,
    n.folds = 10)

  # extract min training error and number of trees to grid
  hyper_grid$optimal_trees[i] <- length(gbm.tune$trees)
  hyper_grid$mean_dev[i] <- gbm.tune$cv.statistics$deviance.mean
  hyper_grid$std_dev[i] <- gbm.tune$cv.statistics$deviance.se
}

#Build selected models in 'gbm' package
set.seed(1234)
mod<-gbm(FlowDiversity~.,data=data[,c(2,3:38)], distribution="gaussian",
       cv.folds=10, keep.data=T, verbose=T, interaction.depth=7,
       shrinkage=0.01, bag.fraction=0.75, n.trees=3400)

#Model stats - R squared, Root Mean Squared Error and deviance
## train: model prediction tested between observed and final model predictions
## cv: model prediction tested between observed and data held-back in
## cross-validation

BRT.stats<-function(model, dat, x){
  output<-data.frame(
    habitat.index=colnames(dat)[x],
    train.R2=(cor(model$fit,dat[,x]))^2,
    train.rmse=caret::RMSE(model$fit,dat[,x]),
    train.dev=calc.deviance(model$fit,dat[,x],calc.mean=T, family="gaussian"),
    cv.R2=(cor(model$cv.fitted,dat[,x]))^2,
    cv.rmse=caret::RMSE(model$cv.fitted,dat[,x]),
    cv.dev=calc.deviance(model$cv.fitted,dat[,x],calc.mean=T,
           family="gaussian"))
}

stats<-BRT.stats(model=mod, dat=data, x=3)
```

Appendix 6D. PCA loadings

Loadings of each control variable in the PCA analysis (**Figure 6.4a** and **6.5**) and the percentage variance explained by the top three principal components (PCs) (**Table 6D**). Only two PCs are reported for Reach PCA (**Figure 6.5a**) as the third has an eigenvalue <1.

Table 6D. *PC loadings for (i) full PCA including all variables, (ii) reach PCA including only reach-level variables and (iii) non-reach PCA including variables above reach-level. Strongest positive (yellow) and negative (blue) variable loadings for each PC highlighted.*

	(i) Full PCA			(ii) Reach PCA		(iii) Non-reach PCA		
	PC1	PC2	PC3	PC1	PC2	PC1	PC2	PC3
Variance (%)	***12.14***	***9.49***	***6.38***	***30.31***	***21.38***	***12.20***	***9.78***	***7.52***
Rch.Confluence	-0.08	0.03	-0.07	0.05	**-0.45**			
Rch.Area	-0.05	-0.18	-0.06	0.31	**0.52**			
Rch.Slope	**0.27**	0.03	0.03	**-0.47**	0.15			
Rch.Order	0.03	**-0.33**	0.01	0.32	**0.60**			
Rch.Elevation	**0.38**	0.17	0.08	**-0.64**	0.22			
Rch.Distance	0.15	0.13	0.10	**-0.41**	0.31			
Rel.Shreve	0.03	**-0.34**	-0.02			0.13	**-0.22**	0.00
Rel.Arable	-0.18	-0.17	0.00			-0.14	**-0.27**	-0.06
Rel.Chalk	-0.11	-0.02	-0.10			-0.12	-0.07	-0.14
Rel.Hard	0.14	-0.12	-0.08			**0.21**	-0.06	-0.04
Rel.ImpGrass	0.04	**-0.24**	0.06			0.11	**-0.26**	0.03
Rel.Mount	**0.30**	-0.15	0.06			**0.37**	0.00	0.11
Rel.Lime	0.12	**-0.21**	-0.07			0.19	-0.16	-0.10
Rel.Sed	-0.02	**-0.31**	**-0.23**			0.09	**-0.32**	**-0.30**
Rel.Sand	-0.02	**-0.32**	**-0.26**			0.09	**-0.34**	**-0.34**
Rel.NatGrass	**0.32**	-0.11	-0.02			**0.39**	0.04	0.05
Rel.Urban	-0.13	-0.09	0.02			-0.11	-0.15	-0.04
Rel.Wood	0.03	-0.15	-0.13			0.10	-0.16	-0.16
Rel.Area	-0.05	**-0.22**	**0.44**			-0.01	**-0.34**	**0.39**
Rel.Slope	-0.08	-0.18	**0.33**			-0.05	**-0.29**	**0.27**
Rel.Power	-0.06	**-0.24**	**0.45**			-0.01	**-0.37**	**0.39**
Ups.Arable	**-0.31**	0.12	-0.06			**-0.36**	0.02	-0.07
Ups.Chalk	-0.17	0.16	-0.14			-0.22	0.11	-0.14
Ups.Hard	**0.22**	0.03	-0.06			0.24	0.13	0.01
Ups.ImpGrass	0.05	-0.15	0.11			0.10	-0.14	0.09
Ups.Mount	**0.31**	0.07	0.09			**0.27**	0.17	0.12
Ups.Lime	**0.21**	-0.05	-0.02			**0.22**	0.01	-0.02
Ups.Sed	-0.09	0.01	**0.35**			-0.12	-0.02	**0.35**
Ups.Sand	-0.02	-0.18	**-0.31**			0.05	-0.20	**-0.37**
Ups.NatGrass	**0.32**	0.00	-0.06			**0.35**	0.15	0.02
Ups.Urban	-0.12	-0.02	0.02			-0.12	-0.08	-0.02
Ups.Wood	0.04	-0.07	-0.16			0.09	-0.08	-0.18
Net.D.Density	0.06	**0.21**	0.05			-0.01	0.15	0.02
Net.E.Density	-0.01	0.05	0.06			-0.03	0.02	0.03

Appendix 6E. Matrices of interaction strength between variables in the BRT model for each habitat index.

Figure 6Ea. *Interaction strength for the* **flow diversity** *BRT. Values are relative within the model. Colour indicates strength of interaction.*

	Rch.Confluence	Rch.Area	Rch.Slope	Rch.Order	Rch.Elevation	Rch.Distance	Rel.Shreve	Rel.Arable	Rel.Chalk	Rel.Hard	Rel.ImpGrass	Rel.Mount	Rel.Lime	Rel.Sed	Rel.Sand	Rel.NatGrass	Rel.Urban	Rel.Wood	Rel.Area	Rel.Slope	Rel.Power	Ups.Arable	Ups.Chalk	Ups.Hard	Ups.ImpGrass	Ups.Mount	Ups.Lime	Ups.Sed	Ups.Sand	Ups.NatGrass	Ups.Urban	Ups.Wood	Net.D.Density	Net.E.Density
Rch.Area	0																																	
Rch.Slope	0.01	0																																
Rch.Order	0.01	0	0.01																															
Rch.Elevation	0.01	0	0	0																														
Rch.Distance	0	0	0.01	0	0.11																													
Rel.Shreve	0	0	0	0	0	0.01																												
Rel.Arable	0	0	0	0	0	0.01	0																											
Rel.Chalk	0	0	0	0	0	0	0	0.01																										
Rel.Hard	0	0	0	0	0	0	0	0	0																									
Rel.ImpGrass	0	0	0	0	0	0	0	0.01	0	0																								
Rel.Mount	0	0	0.01	0	0	0.01	0	0	0	0	0																							
Rel.Lime	0	0	0	0	0	0.04	0	0	0	0	0	0																						
Rel.Sed	0.02	0	0	0	0	0	0	0	0	0	0	0	0																					
Rel.Sand	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0																				
Rel.NatGrass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																			
Rel.Urban	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
Rel.Wood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																	
Rel.Area	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																
Rel.Slope	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.04	0	0	0															
Rel.Power	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0														
Ups.Arable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0													
Ups.Chalk	0	0.01	0	0	0.02	0.01	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0												
Ups.Hard	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0											
Ups.ImpGrass	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0										
Ups.Mount	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
Ups.Lime	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
Ups.Sed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
Ups.Sand	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
Ups.NatGrass	0	0	0	0.01	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
Ups.Urban	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0.04	0.01	0	0	0	0	0	0	0	0	0				
Ups.Wood	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Net.D.Density	0.01	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0.01	0		
Net.E.Density	0.01	0	0	0	0.01	0	0	0	0.02	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.02	0	0	0	0	
Waterbody	0.01	0	0.01	0.03	0.03	0.01	0.03	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0

Figure 6Eb. *Interaction strength for the* **sediment diversity** *BRT. Values are relative within the model. Colour indicates strength of interaction.*

	Rch.Confluence	Rch.Area	Rch.Slope	Rch.Order	Rch.Elevation	Rch.Distance	Rel.Shreve	Rel.Arable	Rel.Chalk	Rel.Hard	Rel.ImpGrass	Rel.Mount	Rel.Lime	Rel.Sed	Rel.Sand	Rel.NatGrass	Rel.Urban	Rel.Wood	Rel.Area	Rel.Slope	Rel.Power	Ups.Arable	Ups.Chalk	Ups.Hard	Ups.ImpGrass	Ups.Mount	Ups.Lime	Ups.Sed	Ups.Sand	Ups.NatGrass	Ups.Urban	Ups.Wood	Net.D.Density	Net.E.Density
Rch.Area	0																																	
Rch.Slope	0	0.01																																
Rch.Order	0	0	0.01																															
Rch.Elevation	0	0.01	0.01	0																														
Rch.Distance	0.01	0	0.02	0	0.02																													
Rel.Shreve	0	0	0	0	0	0																												
Rel.Arable	0	0	0	0	0	0	0																											
Rel.Chalk	0	0	0.01	0.01	0	0	0	0																										
Rel.Hard	0	0	0.01	0	0.01	0	0	0	0																									
Rel.ImpGrass	0.01	0	0	0	0	0	0	0	0	0																								
Rel.Mount	0	0	0	0	0	0	0	0	0	0	0																							
Rel.Lime	0	0	0	0	0	0.01	0	0	0	0	0	0																						
Rel.Sed	0	0	0.02	0	0	0	0	0	0.01	0	0	0	0																					
Rel.Sand	0	0	0	0	0	0	0	0	0	0	0	0	0	0																				
Rel.NatGrass	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0																			
Rel.Urban	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0																		
Rel.Wood	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0																	
Rel.Area	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																
Rel.Slope	0	0	0.01	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0.01	0															
Rel.Power	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0														
Ups.Arable	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0													
Ups.Chalk	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0												
Ups.Hard	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0											
Ups.ImpGrass	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0										
Ups.Mount	0	0	0	0	0.01	0.01	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0									
Ups.Lime	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
Ups.Sed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0							
Ups.Sand	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						
Ups.NatGrass	0	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
Ups.Urban	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
Ups.Wood	0	0	0.01	0	0	0.01	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
Net.D.Density	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0		
Net.E.Density	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Waterbody	0	0	0.02	0	0.02	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0.02	0	0	0	0	0	0	0	0	0.01	0	0	0	0	0

Figure 6Ec. Interaction strength for the **flow type speed** *BRT. Values are relative within the model. Colour indicates strength of interaction.*

	Rch.Confluence	Rch.Area	Rch.Slope	Rch.Order	Rch.Elevation	Rch.Distance	Rel.Shreve	Rel.Arable	Rel.Chalk	Rel.Hard	Rel.ImpGrass	Rel.Mount	Rel.Lime	Rel.Sed	Rel.Sand	Rel.NatGrass	Rel.Urban	Rel.Wood	Rel.Area	Rel.Slope	Rel.Power	Ups.Arable	Ups.Chalk	Ups.Hard	Ups.ImpGrass	Ups.Mount	Ups.Lime	Ups.Sed	Ups.Sand	Ups.NatGrass	Ups.Urban	Ups.Wood	Net.D.Density	Net.E.Density
Rch.Area	0.16																																	
Rch.Slope	0.23	0.12																																
Rch.Order	0.56	0.04	0.08																															
Rch.Elevation	3.1	0.02	2.01	0.23																														
Rch.Distance	0.25	0.07	1.84	0.19	3.87																													
Rel.Shreve	0.01	0.01	0.04	0.01	0.18	0.19																												
Rel.Arable	0.03	0	0.1	0.01	0.04	0.01	0																											
Rel.Chalk	0.02	0	0	0.02	0.04	0.03	0	0.04																										
Rel.Hard	0	0	0	0	0	0	0	0	0																									
Rel.ImpGrass	0.02	0	0	0	0.02	0.05	0	0	0	0																								
Rel.Mount	0.39	0	0.14	0	0.16	0.07	0	0	0	0	0.12																							
Rel.Lime	0	0.05	0.2	0.02	0.42	0.08	0.01	0	0	0	0	0.02																						
Rel.Sed	0.08	0	0.04	0.01	0.02	0.1	0.01	0.01	0.01	0	0	0	0.01																					
Rel.Sand	0	0	0	0	0	0.01	0	0	0	0	0	0	0.01	0																				
Rel.NatGrass	0.01	0	0	0.04	0.01	0.01	0.08	0	0	0	0.01	0.04	0	0.01	0.01																			
Rel.Urban	0.04	0	0	0	0.05	0.02	0	0	0	0	0	0	0	0	0	0.01																		
Rel.Wood	0.18	0.01	0	0.01	0.08	0.01	0.02	0.01	0	0	0.01	0	0.13	0	0	0	0																	
Rel.Area	0.03	0	0.02	0	0.2	0.03	0.01	0	0	0	0	0	0	0	0.01	0.01	0	0																
Rel.Slope	0.04	0	0.24	0.01	0.53	0.02	0.01	0.01	0.02	0	0	0.01	0.6	0.01	0	0.01	0	0.01	0															
Rel.Power	0.13	0	0.01	0.03	0.63	0.12	0	0.01	0	0	0.01	0	0.06	0	0	0	0.01	0.02	0.01	0.05														
Ups.Arable	0.51	0.05	0.03	0.04	0.12	0.5	0	0.57	0	0	0.06	0	0.01	0	0	0.01	0	0.03	0.03	0.05	0.02													
Ups.Chalk	0.07	0	0.1	0	1.7	1.08	0	0.01	0.01	0	0	0	0	0.01	0	0	0	0	0	0.01	0.03	0.01												
Ups.Hard	0.32	0	0.01	0	0.09	0.53	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0.04	0											
Ups.ImpGrass	0.04	0	0.02	0.01	0.07	0.16	0.01	0.01	0	0	0.04	0	0.05	0.07	0.01	0	0	0.02	0	0.09	0.01	0.01	0	0										
Ups.Mount	0.85	0	0.05	0.01	0.05	0.07	0.01	0	0	0	0.15	0.41	0.01	0.01	0.15	0.01	0	0	0	0.31	0	0.01	0	0.01	0									
Ups.Lime	0.01	0	0.03	0	0.2	0.14	0	0	0	0	0	0	0.01	0	0	0	0	0	0	0.01	0	0	0	0	0.01	0.01								
Ups.Sed	0.06	0	0.15	0	0.23	0.03	0.01	0.08	0.01	0	0	0	0.01	0	0	0	0	0.04	0	0	0	0.01	0	0	0.01	0	0							
Ups.Sand	0.1	0.01	0	0.01	0.01	1.47	0	0.01	0	0	0.01	0	0	0	0	0	0	0	0	0	0.03	0.01	0	0	0.01	0	0	0						
Ups.NatGrass	0.02	0	0.01	0	0.04	0.03	0	0	0	0	0	0	0.71	0	0.06	0	0	0	0.01	0.03	0.15	0.12	0	0	0	0.02	0	0.04	0					
Ups.Urban	0.01	0	0	0	0.02	0.01	0	0	0	0	0	0	0	0	0	0	0.03	0	0	0	0.03	0.01	0	0	0.02	0	0	0.01	0	0				
Ups.Wood	0.28	0	0.17	0.05	0.64	0.34	0	0.05	0	0	0.02	0	0.02	0	0	0	0	0.01	0.01	0.11	0	0.18	0.03	0.01	1.54	0.01	0.01	0.02	0.19	0	0			
Net.D.Density	0.59	0	0.55	0.02	0.67	1.47	0.01	0.02	0.02	0	0.03	0.05	0.05	0.01	0.02	0.01	0	0.03	0.01	0	0.07	0.12	0	0	0.02	0.16	0.01	0.01	0	0.01	0.01	1		
Net.E.Density	0.27	0.01	0.18	0.02	0.4	0.82	0.04	0.02	0.01	0	0.03	7.84	0.07	0	0	0.01	0.01	0.14	0	0.02	0.01	0.12	0.26	0.06	0.11	0.51	0	0.02	0.01	0.03	0.16	0.06	0.32	
Waterbody	1.4	0.01	2.59	0.09	2.52	0.37	0	0.03	0.01	0	0.01	0.01	0.32	0.02	0.03	0.04	0.04	0.07	0	0.01	0.01	0.99	0.23	0.28	0.25	0.21	0.09	0.02	0.23	0.19	0.1	0.91	0.1	0.43

Figure 6Ed. *Interaction strength for the* **sediment size** *BRT. Values are relative within the model. Colour indicates strength of interaction.*

	Rch.Confluence	Rch.Area	Rch.Slope	Rch.Order	Rch.Elevation	Rch.Distance	Rel.Shreve	Rel.Arable	Rel.Chalk	Rel.Hard	Rel.ImpGrass	Rel.Mount	Rel.Lime	Rel.Sed	Rel.Sand	Rel.NatGrass	Rel.Urban	Rel.Wood	Rel.Area	Rel.Slope	Rel.Power	Ups.Arable	Ups.Chalk	Ups.Hard	Ups.ImpGrass	Ups.Mount	Ups.Lime	Ups.Sed	Ups.Sand	Ups.NatGrass	Ups.Urban	Ups.Wood	Net.D.Density	Net.E.Density
Rch.Area	0.01																																	
Rch.Slope	1.25	0.57																																
Rch.Order	0.09	0.59	0.97																															
Rch.Elevation	0.45	1.25	0.76	1.94																														
Rch.Distance	1.66	5.69	3.07	1.43	11.4																													
Rel.Shreve	0.1	0.77	0.81	0.18	0.49	0.16																												
Rel.Arable	0.21	0.18	0.41	0.02	0.07	1.88	0.13																											
Rel.Chalk	0.33	0.08	1.92	0	0.59	0.15	0.02	0.01																										
Rel.Hard	0.01	0.01	0	0.02	0.01	0	0.26	0	0																									
Rel.ImpGrass	0.02	0.12	1.4	0.01	0.62	0.37	0.08	0.05	0.64	0.07																								
Rel.Mount	0	0	0.02	0	0.02	0.01	0.09	0.14	0	0.01	0																							
Rel.Lime	0.03	0.02	0.33	0.08	0.29	0.09	0.06	0.19	0	0	0.02	0																						
Rel.Sed	0.17	0.02	0.01	0.01	0.38	0.23	0	0.02	0.1	0.01	0	0	0.02																					
Rel.Sand	0.05	0.05	0.06	0.96	0.52	0.09	0.11	0.06	0	0	0.04	0.02	0.02	0																				
Rel.NatGrass	0	0.01	0.01	0.08	0.02	0.01	0.24	0.03	0	0	0.13	0	0	0.01	0.03																			
Rel.Urban	0.33	0.08	0.01	0.45	0.12	0.21	0.1	0.2	0.05	0	0.7	0	0	0.13	0.04	0.01																		
Rel.Wood	0.24	0.12	0.04	0.05	0.49	0.26	0.16	0.02	0.8	0	0.04	0	0	0.09	0.53	0	0.01																	
Rel.Area	0.09	0.01	0.12	0.13	0.26	0.16	0.04	0.01	0.03	0	0.2	0	0.02	0.02	0.14	0	0	0.13																
Rel.Slope	0.14	0.05	1.94	0.02	0.15	0.96	0.75	0.6	0.14	0	0.03	0	0.02	0.16	0.07	0	0.14	0.14	1.26															
Rel.Power	0.53	0.17	0.36	0.13	0.48	0.19	0.07	0.08	0.01	0	0.05	0	0	0.11	0.32	0	0.21	0.03	1.18	0.82														
Ups.Arable	0.03	0.32	0.37	0.19	2.04	1.31	0.7	0.65	0.18	0	11.9	0.01	0.25	0.18	0.25	1.14	0.07	0.88	0.61	0.56	0.23													
Ups.Chalk	0.06	0.01	0.27	0.07	0.63	2.58	0.01	0.01	0.08	0	0	0	0	0	0	0	0	0	0	0	0	0.01												
Ups.Hard	0.03	0.08	0.06	0	0.25	0.06	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0.04	0.02	0											
Ups.ImpGrass	0.12	0.04	0.74	0.02	1.15	11.5	0.3	0.08	0.15	0	1.63	0.03	0	0.06	0.02	0	0.15	0.61	0.05	0.59	0.01	0.38	0.03	0										
Ups.Mount	0.01	0	0.41	0.05	0.49	2.28	0.13	0.05	0	0	0	0.03	0	0.02	0.05	0	0	0.01	0	0.05	0	0.09	0	0	0.01									
Ups.Lime	0	0.06	0.04	0.16	0.31	0.52	0.15	0.04	0.01	0	0	0	0	0.02	0.03	0	0.03	0	0.01	0.15	0.03	0.11	0.01	0	0	0								
Ups.Sed	0.74	0.18	0.05	0.11	0.65	1.87	0.62	0.49	0.2	0.02	0.04	0	0.22	0.54	0.24	0.03	0.24	0.35	0.01	0.23	0.14	0.06	0	2.31	0	0.01	12.7							
Ups.Sand	0.1	0.11	0.11	0.13	1.45	0.69	0.81	0.25	0	0.01	0.49	0.04	0	1.63	1.7	0	0.02	0.03	0.02	0.13	0.14	0.14	0.01	0	0.08	0	0	3.14						
Ups.NatGrass	0	0	0.03	0.42	0.46	0.29	0.05	0	0	0	0.02	0	0	0.01	0.02	0.05	0.02	0.01	0	0.01	0.04	0.6	0	0	0.01	0.08	0	0.01	0					
Ups.Urban	0.12	0.05	0.23	0.1	0.13	0.18	1.59	0.08	0.01	0	0.05	0.02	0.04	0.01	0.06	0	0.69	0.5	0.03	0.21	0.02	0.17	0.02	0	0.17	0	1.51	0.18	0.06	0				
Ups.Wood	0.46	0.11	0.02	0.03	1.14	1.43	1.02	0.29	0.04	0	0.03	0	0	0.01	0.02	0	0	6.46	0	0.05	0.01	0.44	0	0	0.1	0	0.02	0.04	0.08	0.01	0.39			
Net.D.Density	3.76	0.4	0.47	0.58	6.95	1.15	0.55	1.18	0.03	0.01	1.1	0.02	0.04	0.17	0.32	0.02	0.08	0.15	0.05	1.91	0.47	1.79	0.19	0	0.24	0.04	0.65	2.54	0.17	1.98	0.2	0.01		
Net.E.Density	0.64	0.44	0.85	0.38	0.66	1.72	0.1	1.14	0.08	0	0.07	0.01	0.07	0.4	1.14	0	0.4	0.85	0.14	0.45	1.08	0.19	0.1	0.01	0.73	0.01	0.01	1.6	0.15	0.01	0.68	0.08	0.52	
Waterbody	0.13	1.64	5.76	0.9	8.03	6.15	3.96	0.43	0.01	0	0.37	0.21	0.1	0.04	0.29	0.03	0.17	0.77	0.12	0.67	0.19	15.9	0.17	0.17	2.39	1.82	1.56	1.2	2.39	2.16	1.11	3.92	2.44	0.62

END