Capitalising on the Big Data era: establishing a multi-source monitoring framework for England's natural capital assets and flows

Submitted for the degree of Doctor of Philosophy

School of Archaeology, Geography and Environmental Science, University of Reading

Merry R. D. Crowson Davidson

June 2024

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Merry R. D. Crowson Davidson

4th of June 2024

Acknowledgements

I am grateful to my supervisors for their support and input over the years – Nathalie Pettorelli, Andrew Wade, Nick Isaac and Ken Norris. I am very lucky to have been exposed to such a wide range of perspectives and ways of working during my PhD.

I am deeply grateful to Nathalie Pettorelli for all her mentoring, advice, encouragement, and feedback, and for being a formidable ally over the past 6 years. Many thanks to Andrew Wade for helping shape my ideas and keep my PhD on course, and for his ongoing support with the water quality work. Many thanks to Nick Isaac for all the conversations and great ideas - they were often what got things across the line, and always provided motivation to keep going. Many thanks to Ken Norris for providing perspective and feedback on the big picture - it was always great to catch up.

Many thanks to everybody at the Institute of Zoology who has provided mentoring and good humour over the years. A special thanks to Henny, Henry and Jake who have been the best officemates I could have hoped for in the final year of my PhD. Many thanks also to The Opportunists for welcoming me to their online meetings during the pandemic.

My PhD was generously funded by NERC (grant number NE/R012229/1), and I received invaluable support and training from the QMEE Doctoral Training Programme at the University of Reading. I would like to thank the Greenspace Information for Greater London CIC for enabling access to data on green spaces in Greater London for the work in Chapter 4, and for taking the time to share their experience and knowledge of the datasets.

Finally, many thanks to my family. Fabian, I would not have been able to start my PhD without your support and enthusiasm, let alone finish it. Max and Hannah, you have been flying colours of red and lime green and yellow, the perfect distraction from my work. Many thanks to Daisy for all the great conversations about research, and to mum and dad, who did a lot of childcare with a lot of grace. Many thanks to my extended family for always asking me "how is the PhD going?" and, importantly, "what are you going to do next?".

Abstract

The past decade has seen a growth of natural capital accounting both internationally and nationally. The term natural capital refers to the elements of nature that directly and indirectly produce value or benefits to people, including ecosystems, species, freshwater, land, minerals, the air and oceans, as well as natural processes and functions (Natural Capital Committee 2014). As an approach it emphasises the process of valuation, namely estimating the relative importance, worth, or usefulness of natural capital to people, typically to enable better governance. In this thesis, I explore the potential of big data and associated techniques to operationalise the natural capital framework at a national scale in England, through a better understanding of the relationship between natural capital assets and the benefits that flow from them. I take an interdisciplinary approach, using the literature review in Chapter 1 to identify key gaps in the state of the art, and addressing these gaps in the following chapters, finishing with a discussion of the implications of these findings in the final chapter (Chapter 5). The results from this thesis demonstrate how diverse and emerging environmental datasets can capture important aspects of sociocultural value that are otherwise hard to include in a formal valuation process (Chapter 2), enable spatially targeted management (Chapter 3), and facilitate natural capital monitoring (Chapter 4). In Chapter 2, I demonstrate the potential of crowdsourced data to capture the sociocultural value of designated areas and show that species richness has a significant positive effect on public interest in designated areas. In Chapter 3, I show that population density is a driver of the relative importance of agricultural land use as a source of N and P in river catchments in England. In Chapter 3, I demonstrate that significant dependency exists between the quantity, quality and spatial configuration of green spaces in London, and that there is potential to maintain highly biodiverse areas in cities, without assigning large areas to this. Taken together, these results realise some of the potential of the big data era to support the natural capital framework and its implementation, as well as pointing to some of the limitations of this approach.

Table of Contents

List of Abbreviations

AIC Akaike's Information Criterion

AONBs Areas of Outstanding Natural Beauty

API Application programming interface

AUC Area under the curve

BFIHOST Base flow index based on the Hydrology of Soil Types classification

CAI_AM Area weighted mean core area index

CEH Centre for Ecology and Hydrology

Defra Department for Environment, Food and Rural Affairs

GIS Geographic Information System

glm generalised linear model

IHDTM Integrated Hydrological Digital Terrain Model

IPBES Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services

LNRs Local Nature Reserves

LSI Landscape shape index

MEA Millennium Ecosystem Assessment

NDVI Normalized difference vegetation index

NNRs National Nature Reserves

N Nitrogen

P Phosphorus

RMSE Root mean square error

ROC Receiver-operating characteristic curve

RP Reactive Phosphorus

SACs Special Areas of Conservation

SPAs Special Protection Areas

SSSIs Sites of Special Scientific Interest

TON Total Oxidised Nitrogen

UNEP The United Nations Environment Programme

WWTWs Waste water treatment works

Publications and conference presentations

Included in the thesis:

Crowson, M., Isaac, N. J. B., Wade, A. J., Norris, K., Freeman, R., and Pettorelli, N. (2023). Using geotagged crowdsourced data to assess the diverse socio-cultural values of conservation areas: England as a case study. Ecology and Society, 28(4):28. https://doi.org/10.5751/ ES-14330-280428

Conference presentations:

Crowson, M., Isaac, N. J. B., Wade, A. J., Norris, K., Freeman, R., and Pettorelli, N. (2023). Using geotagged crowdsourced data to assess the diverse sociocultural values of conservation areas: England as a case study. British Ecological Society Annual Meeting, Belfast, December 2023. Oral Presentation.

Crowson, M., Isaac, N. J. B., Wade, A. J., Norris, K., Freeman, R., and Pettorelli, N. (2023). Using geotagged crowdsourced data to assess the diverse sociocultural values of conservation areas: England as a case study. British Ecological Society Macroecology Conference, Birmingham, July 2023. Oral Presentation.

Publications complementary to the thesis:

Pettorelli, N., Williams, J., Schulte to Bühne, H., and **Crowson, M.** (2024) Deep learning and satellite remote sensing for biodiversity monitoring and conservation. Remote Sensing in Ecology and Conservation. https://doi.org/10.1002/rse2.415

Crowson, M. †, Williams, J. †, Sharma, J., and Pettorelli, N. (2024). Challenges for monitoring the growth of artificial turf with remote sensing. Environmental Monitoring and Assessment, 196, 580. https://doi.org/10.1007/s10661-024-12724-w.

Elphick, A., Ockendon, N., Aliacar, S., **Crowson, M.**, and Pettorelli, N. (2024). Long-term vegetation trajectories to inform nature recovery strategies: The Greater Côa Valley as a case study. The Journal of Environmental Management. https://doi.org/10.1016/j.jenvman.2024.120413.

Chapter 1: Introduction and literature review

1. Introduction

Human activity is leading to global biodiversity loss and increasing pressure on natural resources, compromising the ability of the planet's natural environment to sustain future generations (Watson et al. 2005; FAO 2019; IPBES 2019). A growing recognition that environmental degradation has a negative impact on human wellbeing has shifted the framing and purpose of conservation towards balancing the requirements of both biodiversity and people, recognising the importance of sustainable and resilient interactions between human societies and the natural environment (Mace 2014), and looking for synergies between the development and conservation agendas (Allam and Naser 2018). The United Nations Environment Programme (UNEP) Millennium Ecosystem Assessment (MEA), which ran between 2001 and 2005, focused global efforts on assessing the consequences of ecosystem change for human well-being and the scientific basis for action needed to enhance the conservation and sustainable use of those systems (Watson et al. 2005). The recognition within the MEA of the economic contribution of nature to human well-being has been very influential (Mace 2014), as has a focus on ecosystem services – defined in the MEA as the benefits people obtain from ecosystems (Watson et al. 2005). These benefits are commonly divided into provisioning services such as food, water and timber; regulating services such as floods and disease control; cultural services that provide recreational, aesthetic, and spiritual benefits; and supporting services such as soil formation and photosynthesis (Watson et al. 2005).

The focus on a sustainable provision of ecosystem services has led to the emergence of natural capital accounting, aimed at measuring and reporting on natural capital and the flow of benefits we receive from it in a systematic way (SEEA 2023). The term natural capital refers to the elements of nature that directly and indirectly produce value or benefits to people, including ecosystems, species, freshwater, land, minerals, the air and oceans, as well as natural processes and functions (Natural Capital Committee 2014a). In England the concept of natural capital became central to environmental discourse when the government pledged to take a natural capital approach in its 25 Year Environment Plan (Defra 2018a). This framework necessitates regular monitoring of the nation's natural capital assets to ensure efficient management of resources, to maximise the benefits we receive from nature and ensure that the integrity of our natural capital is not compromised for future generations (Defra 2018a).

Monitoring natural capital at a national scale is a huge challenge because it means collecting, storing, and processing large datasets, overcoming data gaps, and developing suitable indicators for the status of natural capital and the benefits we receive from it (Defra 2018b, Natural England 2018). Big data offers new opportunities to face this challenge. Big data broadly refers to the increasing volume, variety, and velocity of data streams over the past 20 years or so (Hampton et al. 2013; Chen et al. 2014). Developments in this area make it possible to harness new types of information, such as remote sensing data (Pettorelli et al. 2016) and crowdsourced data (Isaac et al. 2014); to use increased processing power through platforms such as Google Earth Engine (Gorelick et al. 2017); and to apply new methods developed to process and analyse large and varied datasets, such as machine learning algorithms (Willcock et al. 2018).

This PhD aims to explore the potential of big data to increase our understanding of the relationship between natural capital assets and the benefits that flow from them, taking an interdisciplinary approach, and through this contribute to the operationalisation of the natural capital framework at a national scale. This first chapter combines the introduction with a literature review, to set the context for the work overall and define key terms and concepts, as the results chapters (Chapters 2, 3 and 4) are presented in the format of papers, and thus include the relevant literature for each topic. The literature review will start (in Section 1.2.1) with an outline of the policy context that led to the emergence of the natural capital framework, followed in Section 1.2.2 by a comparison of the various frameworks currently used to account for nature and the benefits we derive from it. Section 1.2.3 focuses on the natural capital framework as it is being used in the UK, including an overview of the different conceptual models that are used when implementing the natural capital framework and a description of some of the major work that has been carried out using the natural capital framework nationally. Section 1.2.4 examines the importance of valuing natural capital in decision making, and the central role of understanding the link between natural capital assets and benefits that flow from them in the valuation process. Section 1.2.5 examines natural capital as a big data challenge, including the potential of new datasets, methods and platforms to contribute to monitoring efforts. Finally, Section 1.2.6 lays out the research questions my thesis aims to answer, that target key gaps in knowledge identified in the previous sections.

1.2 Literature review and research objectives

1.2.1 Policy context

The MEA fundamentally changed approaches to policy internationally, moving policy discourse away from a "nature for itself" framing (Mace 2014), and emphasising the specific, quantifiable benefits that society receives from nature (Hungate and Cardinale 2017, Pan and Vira 2019). For example, the EU Biodiversity Strategy to 2020 addressed the need to account for ecosystem services through biophysical mapping and valuation (Maes et al. 2012), and the concept of payment for ecosystem services has become increasingly popular (Farley and Constanza 2010). Payment for ecosystem services describes schemes in which the beneficiaries, or users, of ecosystem services provide payment to the providers of ecosystem services (Defra 2013). In this context, the concept of natural capital has gained traction internationally, sharpening the focus on both the biodiversity and non-biodiversity assets that underpin the ecosystem services we receive from nature, and the need to conserve it for future generations. UNEP produced a global map of natural capital in 2014 (Dickson et al. 2014) and the concept of natural capital features prominently in the European Union's seventh Environment Action Programme 'Living well, within the limits of our planet' (European Parliament and Council 2013) as well as in "The Economics of Biodiversity: The Dasgupta Review" (Dasgupta 2021). Most recently, the 2022 "Assessment Report on Diverse Values and Valuation of Nature", by the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), emphasises the need for decision-makers to understand and account for the wide range of nature's values in policy decisions.

In the UK, the UK National Ecosystem Assessment (UKNEA 2011) was followed by the Government White Paper, 'The Natural Choice: Securing the Value of Nature', which aimed to 'put natural capital at the centre of economic thinking and at the heart of the way we measure economic progress nationally' (Defra 2011, p. 4). To help achieve this, the Natural Capital Committee was set up in England in 2012 and ran until December 2020 as an independent advisory committee to advise the government on the sustainable use of natural capital in England, including helping the government develop its 25 Year Environment Plan for protecting and improving the environment (Natural Capital Committee 2020, Defra 2018a), some of which was made legally binding through the Environment Act 2021 (Environment Act 2021). The Office for National Statistics now publishes national natural capital accounts annually (Office for National Statistics 2023a), as the UK Government continues to take a natural capital approach to environmental policy.

1.2.2 Frameworks to account for nature and the benefits it provides

The recognition within policy of the economic contribution of nature to human well-being has led to a plethora of different frameworks to account for these benefits, including ecosystem services, natural capital and nature's contributions to people.

Within the ecosystem services framework *ecosystems* provide *ecosystem services* to humans. Ecosystems refer to a dynamic complex of plant, animal, and microorganism communities and their non-living environment interacting as a functional unit (UKNEA 2011). Ecosystem services can be defined as the conditions and processes through which natural ecosystems, and the species that make them up, sustain and fulfil human life (Goulder and Kennedy 1997). This includes services such as pollination, purification of air and water, mitigation of floods and droughts, provision of timber, recreational opportunities, and aesthetic beauty (Goulder and Kennedy 1997). N*atural capital* refers to the elements of nature that directly and indirectly produce value or benefits to people, including ecosystems, species, freshwater, land, minerals, the air and oceans, as well as natural processes and functions (Natural Capital Committee 2014a). Natural capital is thus broader than ecosystems, as these elements need not be interacting, as is implicit in the definition of an ecosystem (Mace et al. 2015), and the term includes fossil fuels and minerals (Mace 2019).

The term natural capital has its roots in economics (Missemer 2018) and refers to the economy's environment and natural resource endowment (Barbier 2013). In economics, natural capital sits alongside produced or manufactured capital (roads, machines, buildings) and human capital (health, knowledge, institutions, culture). Together these three types of capital underpin our economy and ultimately human wellbeing (Mace 2019). This shift towards language and concepts taken from economics can be seen an attempt to ensure that nature is properly valued by decision makers and planners, and to engage businesses in conservation efforts. However, critics of the natural capital approach to conservation highlight that the value of nature is infinite and boiling this down to a series of benefits means essentially "selling out" on nature (McCauley 2006, Schröter et al. 2014).

Within the natural capital framework *stocks* of natural capital support a *flow* of benefits to society. This explicit use of language taken from economics is one of the elements that distinguished the natural capital framework from the ecosystem services framework, which have been closely equated (e.g. Kareiva et al. 2011). Within the natural capital framework, the flows from the stocks of natural capital are sometimes referred to as ecosystem services (e.g.

Harrison et al. 2017, Jones et al. 2016), but in other publications they are referred to as benefits, which can include both services and goods (e.g. Mace et al. 2015). These terms will be addressed in more detail in the next section, but generally in the literature the distinction between ecosystem services and benefits is not clear-cut and they often refer to much the same thing. Thus, the natural capital framework is closely related to the ecosystem services framework but uses language from economics to speak more directly to economic planning and development efforts, as well as sharpening the focus on the stocks of natural capital that underpin the benefits humans receive from nature. Thus, most framings of natural capital treat ecosystem services as a subset of natural capital, representing flows of benefits from the stocks.

Another framework that has emerged more recently from the IPBES is *nature's contributions to people* (Díaz et al. 2018). Within this framework nature's contributions to people are all the contributions, both positive and negative, of living nature (diversity of organisms, ecosystems, as well as their associated ecological and evolutionary processes) to people's quality of life (Díaz et al. 2018). This framework was put forward as it was felt that the development of the ecosystem services and natural capital frameworks were dominated by knowledge from the natural sciences and economics, and that both have failed to engage with perspectives from the social sciences (Díaz et al. 2018, Norgaard 2010), or those of local practitioners, especially indigenous peoples (Díaz et al. 2018). However, Braat (2018) strongly refuted these claims, pointing to (i) the fact that more than half of the 650 publications in the journal "Ecosystem Services" in the period 2012–2017 address social aspects and are based on social science methods; (ii) the publication of dozens of articles on cultural ecosystem services in the journal in the same period; and (iii) the small but increasing number of publications on Indigenous Knowledge. In a response to Díaz et al. (2018), de Groot and colleagues (2018) call for nature's contributions to people and ecosystem services to be regarded as synonyms, suggesting that their use may simply differ based on the audiences and purposes.

This plurality of frameworks, and the fact that some terms have multiple definitions, creates some challenges with regards to maintaining scientific cohesion, and threatens to undermine the challenging process of securing international commitments (Peterson et al. 2018). But building practical knowledge for sustainability in a diverse world is likely to require a diversity of approaches, each of which develops with its own history and policy context. The next section follows the development of a conceptual model for natural capital in the UK.

1.2.3 The natural capital framework

1.2.3.1 Scope

The decision of the UK government to take a natural capital approach in the 25 Year Environment Plan (Defra 2018a) has meant that this approach has become central to environmental discourse and policy in England. For the most part, the proposals in the 25 Year Environment Plan apply to England only because environmental policy is devolved and responsibility rests with the Scottish Government, Welsh Government and Northern Ireland Executive. Strictly speaking, since the UK Government is responsible for a number of policies and programmes which affect sectors across the UK and internationally, some aspects of the 25 Year Environment Plan apply to the UK as a whole (Defra 2018a). It is, for example, the UK that is signatory to international treaties such as the UN Convention on Biodiversity, rather than the individual countries that make up the union. However, the uptake of the natural capital framework varies between the countries that make up the union. Significantly, Wales has not explicitly adopted the natural capital framework, instead opting for and using the term "natural resources", and focusing policy on the wellbeing of future generations (Welsh Government 2015). This being said, because the natural capital approach is part of academic discourse to a greater or lesser extent in all of the countries that make up the union, this section looks in detail at the conceptual models and applications for natural capital that are being used in the UK.

1.2.3.2 The development of a conceptual model for natural capital

In the UK, different conceptual models for natural capital have emerged (Harrison et al. 2017). They have some elements in common, specifically the idea of *stocks* of natural capital and *flows* from them that contribute to human wellbeing and livelihood. A stock is the amount of the natural resource (biotic/abiotic) which make up natural capital; a flow relates to the services and benefits arising from natural capital (Harrison et al. 2017). Benefits flow from the stocks of natural capital (e.g. Defra 2018a, Mace et al. 2015; Figure 1), with benefits defined as an advantage or good effect (Harrison et al. 2017). For example, from our stocks of soils, fresh water and species we receive the benefit of food. Likewise, our stock of forests provides the benefits of timber, CO_2 sequestration and flood protection.

Natural capital stocks can be organised into classes called "assets" (Mace et al. 2015), defined as "things of value" (Harrison et al. 2017). The Natural Capital Committee conceptualises natural capital as a series of overlapping assets that are formed and maintained by natural processes (Mace 2019). Some models define the assets they consider in a structured way. For example, Mace and colleagues 2015 considers the following as natural capital assets: species, ecological communities, soils, freshwaters, land, minerals, atmosphere, subsoil assets, oceans, and natural processes and functions. However, even when natural capital assets are listed explicitly in this way, it is often broad habitat classes that are taken as the "accountancy units" of natural capital, by using the major land use classes (e.g. Natural Capital Committee 2014a, Mace et al. 2015). In other cases, the natural capital assets are not clearly listed within the framework and the categories considered are much more flexible, depending on what is being studies - an asset could be lakes (Harrison et al. 2017), green infrastructure features in urban areas (Office for National Statistics 2023b) or riverine vegetation (Harrison et al. 2017).

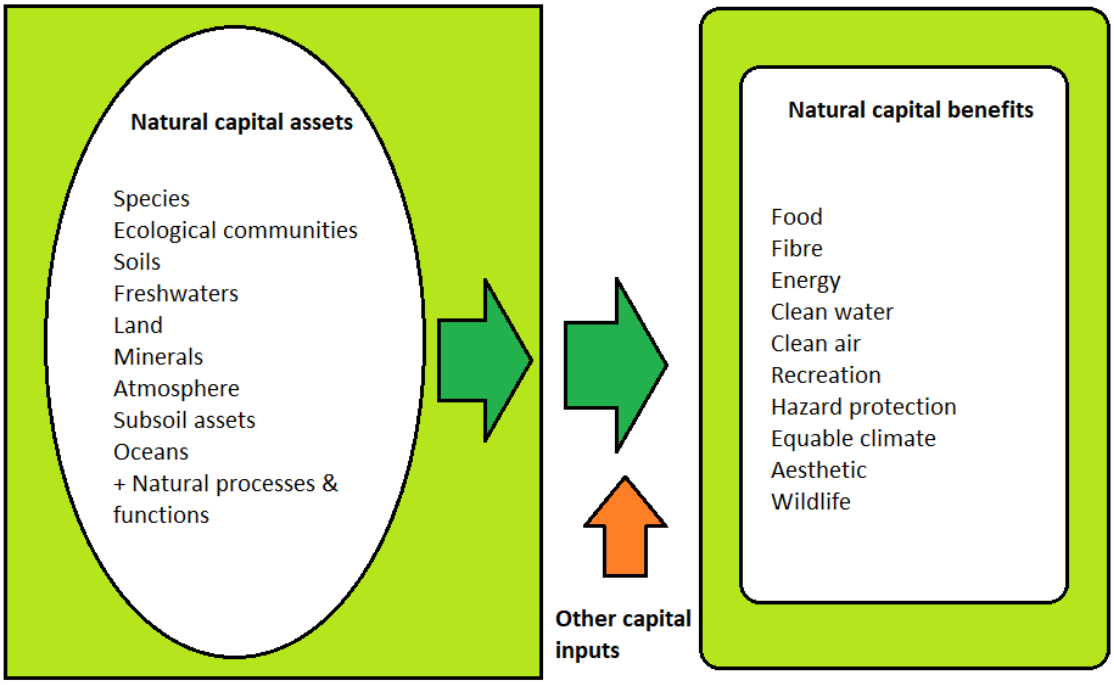


Figure 1: Conceptual model for natural capital, following the simplified framework in Mace et al. 2015.

In some conceptual models of natural capital, the flows from the stocks of natural capital are called ecosystem services rather than benefits (e.g. Jones et al. 2016, Harrison et al. 2017), providing a clear link to the ecosystem services framework. In various other models the flows from natural capital are broken down into ecosystem services, goods and benefits (e.g. Mace et al. 2015, Natural Capital Committee 2014a, Dickie et al. 2014; Figure 2), with other types of capital (human, manufactured) combining with natural capital at various stages. However, the terms ecosystem services, goods and benefits are not defined clearly in these publications so the difference between them and how they lead on from each other is not clear. Presumably, the difference between goods and services relates to the economic terms, with goods relating to the physical objects and services an activity that is performed. However, this difference between services, goods and benefits in the context of natural capital is a bit nebulous and creates challenges when implementing the natural capital framework.

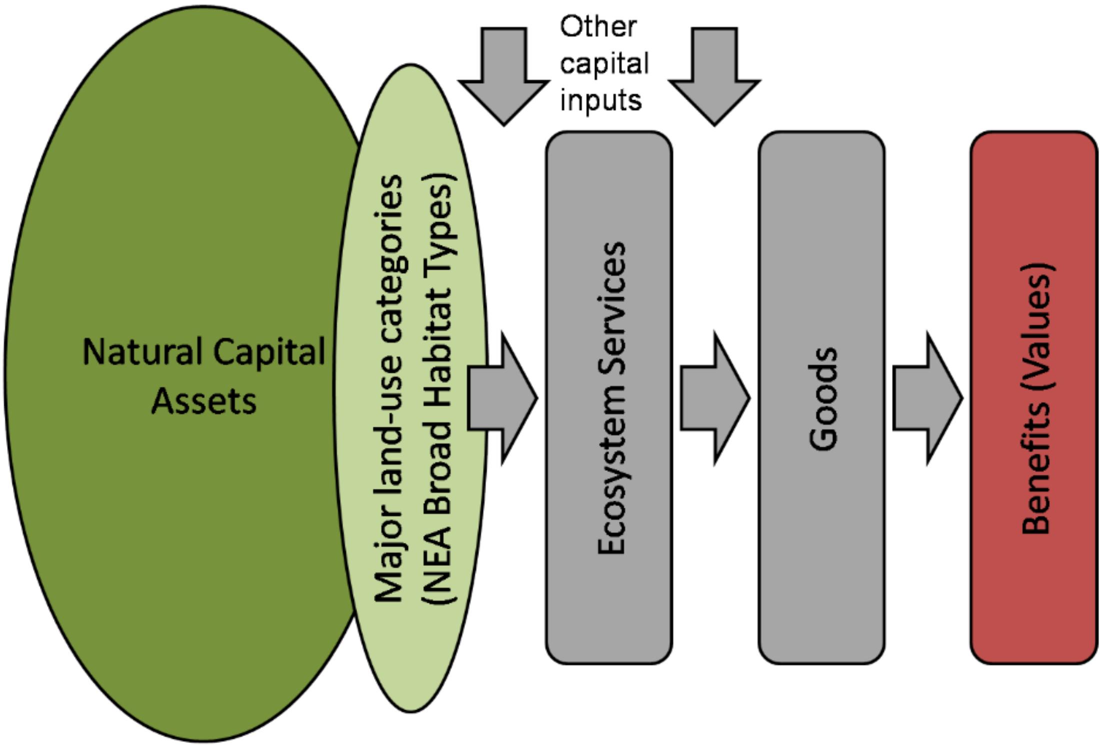


Figure 2: Example of a conceptual model for natural capital in which flows from natural capital are broken down into ecosystem services, goods and benefits. Taken from Natural Capital Committee (2014a).

Whilst the basic building blocks of the natural capital framework are the concepts of stocks of natural capital assets and flows of benefits/ecosystem services, some conceptual models for natural capital include other elements such as pressures on natural capital assets and societies' responses (Harrison et al. 2017). Other models make more refined distinctions, for example between potential supply of ecosystem services and actual supply where user demand exists (Jones et al. 2016).

In this work, a simple natural capital framework will be adopted, with stocks of natural capital assets and flows of benefits from them (Figure 1). This means that benefits and ecosystem services are considered synonyms and used interchangeably throughout. I hope that by using the simple natural capital framework it is possible to maximise overlap with other work and minimise confusion due to differences in terminology.

Natural capital here refers to the elements of nature (living and non-living) that produce value or benefits to people (directly and indirectly), such as the stock of forests, rivers, land, minerals and oceans, as well as the natural processes and functions that underpin them (Natural Capital Committee 2013).

1.2.3.3 Work using the natural capital framework in the UK

A range of work has been carried out using the natural capital framework, in both academic and more applied settings. There is a large body of work that maps or otherwise records natural capital at a particular place and time, as a kind of "stock-take" or "asset-register" of natural capital. An example of this type of work is the Mapping Natural Capital (RP02404) project, a collaboration between the Centre for Ecology and Hydrology (CEH) and Natural England (CEH and Natural England 2017). The aim of this project was to produce a series of England-wide natural capital maps at a 1 km scale, to contribute to our understanding of where our natural capital is. They mapped 10 different aspects of natural capital: soil carbon, soil nitrogen, soil pH, soil phosphorus, soil bacteria, soil invertebrates, headwater stream quality, carbon in vegetation, nectar plant diversity for bees, and plant indicators for habitats in good condition (CEH and Natural England 2017). Many similar natural capital mapping projects have been carried out more locally, for example at a county (e.g. Lear et al. 2021) or city level (e.g. Birmingham City Council 2013, Mayor of London 2017). These types of mapping projects help highlight the benefits provided by the natural environment, and can help planners and decision makers protect and restore the natural environment for the benefit of people. However, they are static, as they simply provide a snapshot of a particular time and place. It is of course possible to repeat these mapping projects with the aim of detecting change, but this is not usually done and thus it is not clear if the methods used in these studies will be sensitive enough to allow change detection in stocks and flows at the magnitude and speed that they are occurring.

The work to develop natural capital accounts for the UK by the Office for National Statistics and the Department for Environment, Food and Rural Affairs (Defra) has a clear focus on methods to map and quantify assets on a regular basis in order to detect changes in stocks and flows. Their aim is to incorporate natural capital into the UK Environmental Accounts, including changes to the extent and condition of the physical environmental assets and to the value of services provided by ecosystems. In addition to annual national capital accounts (Office for National Statistics 2023a), they create habitat accounts for urban areas (Office for National Statistics 2023b), woodland (Office for National Statistics 2022a) and health benefits from recreation (Office for National Statistics 2022b). The natural capital estimates are measured in physical units (for example, how many trees or acreage of forest there are), in terms of their condition (using condition indicators) and in monetary terms (what the value of woodland is) (Bright et al. 2019). Compared to the rest of the work discussed in this section, there is a strong focus here on the final step of valuing natural capital stocks in monetary terms, which they calculate as the net present value of current and future service flows. These final monetary estimates appear low even to those involved in the project (Bright et al. 2019). One reason for this is that the risk and cost of an asset deteriorating beyond some biological threshold and collapsing (a fish population, for example) is not factored into the valuation process, including the costs that would incur during recovery.

There are various studies that do link the status of assets to whether the asset is at risk of failing to provide the benefits that society requires of it, providing a "risk-register" to sit alongside an "asset-register". For example, the risk register in Mace et al. 2015 uses the asset-benefit relationship for ten classes of benefit (food, timber, energy, aesthetics, freshwater quality, recreation, clean air, wildlife, hazard protection and equable climate) across eight broad habitat types in the UK, to assess the status and trends of natural capital assets relative to societal targets. They use existing regulatory limits and policy commitments to allocate a score of high, medium or low risk to the asset-benefit relationships, enabling monitoring efforts and protection to be assigned to those natural assets where benefits are at highest risk (Mace et al. 2015). A combination of an "asset register" and a "risk register" can also be found elsewhere, for example in the Natural Capital Asset Check Approach proposed in the UK National Ecosystem Assessment Follow-on (Dickie et al. 2014), which aims to identify the amount and condition of natural assets, and then make an assessment of their current and future performance, with performance measured in terms of their ability to support human well-being (eftec 2012). More recently, Rees and colleagues (2022) created a natural capital risk register for marine systems in North Devon.

From the beginning, the natural capital approach has attempted to reach beyond those working in a traditional conservation setting and appeal to decision-makers more broadly, including economists and planners. There are a range of tools that have been created to help the public, private and third sector organisations to manage the environment as an asset that delivers benefits for society (Ecosystems Knowledge Network 2019). These are tools that are capable of analysing information and producing an output that can inform decision-making. A widely used tool in this area is InVEST, created by the Natural Capital Project at Stanford University (Sharp et al. 2018). This tool combines a suite of open-source models that map and value a range of ecosystem services, including carbon storage and sequestration, crop pollination, recreation, sediment retention, scenic quality and water purification (Sharp et al. 2018). InVEST models are spatially-explicit and define how changes in an ecosystem's structure and function are likely to affect the flows and values of ecosystem services across a land- or a seascape (Ecosystems Knowledge Network 2019). InVEST enables decision-makers to assess quantified trade-offs associated with alternative management choices and to identify areas where investment in natural capital can enhance human development and conservation. There are other tools designed to assess a range of ecosystem services, such as ARIES (ARtificial Intelligence for Ecosystem Services), EcoServ-GIS, which is a Geographic Information System (GIS) toolkit for mapping ecosystem services at a county or regional scale, and NCRAT (Natural Capital Register and Account Tool), which is a publicly accessible, excel-based natural capital accounting tool (Environment Agency 2023a). In addition to these, there are a range of tools designed with specific assets and services in mind, for example i-Tree Eco, which is designed to quantify the structure and environmental effects of urban forests or trees, and their value to communities (Ecosystems Knowledge Network 2019). Finally, there are also publications that study natural processes within the natural capital framework, whether it be the processes that link stocks and flows, or the processes that control trade-offs and synergies between benefits. An example of this is the work on context dependency by Spake et al. (2019), which addresses the challenge of identifying both why and where management actions are most effective for enhancing natural capital across large geographic areas. The work puts forward a framework to achieve this by creating 'effect maps' across large spatial extents, which quantify how the effects of key drivers of ecosystem responses vary across broad geographic extents. Whilst this work is firmly framed within the natural capital framework, the findings on 'cross-scale interactions' will be of interest more broadly to those working in environmental management or studying landscape ecology. This kind of work thus provides important understanding on natural processes, targeted at the monitoring and implementation needs of the natural capital framework.

1.2.4 Valuing natural capital through an understanding of the asset-benefit relationship

In environmental economics, valuation refers to the process of estimating the relative importance, worth, or usefulness of natural capital to people (Natural Capital Coalition 2016). Critics of the natural capital framework raise concerns about the commodification of nature during the valuation process (Silvertown 2015), and concerns that the process of valuation itself overlooks the intrinsic value of nature (Díaz et al. 2018).

However, others argue that if the benefits provided by nature are not assigned a value, they will, by default, be assigned no value in decision making processes (Mace 2019). In addition, the process of valuation of nature does not necessarily involve monetisation (Constanza et al. 2012). The work by the Office for National Statistics to incorporate natural capital into national accounts, described in the previously in Section 1.2.3.3, does involve putting a final monetary label on natural capital assets and benefits. However, there are a range of other ways in which nature is more commonly valued (Pearson 2016), and this valuation can be qualitative, quantitative, monetary or a combination of these, much in the same way as with ecosystem services. There are, for example, non-monetary techniques that focus on human expressions of preference, which can be studied through various methods including surveys, focus groups, participative mapping and official national statistics (Kelemen et al. 2016). Seeking to value natural capital assets through some kind of natural accounting process is in fact what any natural capital approach aims to achieve, as decision makers are often faced with trade-offs between different assets and questions about where best to invest limited financial resources to maximise benefits. As well as supporting decision making, valuation should be a way for governments, institutions and individuals to take responsibility for the natural capital assets that underpin society, monitor the condition of these assets over time and ensure that their condition do not fall below critical levels (Mace 2019).

A full valuation of our natural environment is challenging, as it underpins every aspect of our life and can thus be considered of infinite value. In addition, the question "of value to whom?" or "of benefit to whom?" quickly appears in relation to valuing natural assets (Ghermandi and Sinclair 2019, Wilkins et al. 2021), as not everybody values the same thing, and most benefits are not distributed equally through space and amongst people. Various approaches to natural capital accounting and valuation have developed in parallel. Mace (2019) outlines some of the existing approaches, including the ecosystem-service based approach, which uses the value of the flows of benefits from assets to assess their value. This is similar to the approach found in the Parliament's POSTnote from December 2016, which considers the value of a natural capital asset as the overall benefit it adds to society (POSTnote 542). In contrast, the ecosystem-capability approach accounts for the fundamental and irreplaceable properties of natural capital through metrics that measure natural capital assets condition in terms of its potential to support these functions without interventions (Mace 2019). Whatever approach is taken, using metrics to account for the full value of assets is challenging, as a range of different metrics are needed to reflect the multiple benefits received from natural capital assets, and many of these metrics still need to be developed to operationalise the approach.

Understanding the relationship between assets and the benefits they provide is central to valuation and natural capital accounting, in order to map certain benefits to particular assets (Harrison et al. 2017) or to understand the relationship between the condition of an asset and the value of the benefit provided to people (Mace et al. 2015). When relating assets to benefits it is common to assume that a single asset provides a particular benefit (e.g. Mace et al. 2015, Natural England 2018). However, this is a simplification as it is often the case that multiple assets come together to provide benefits. For example, no single assets can provide us with the benefit of drinking water, and instead it is provided by a range of different assets coming together at a landscape scale. Indeed, multiple assets are combining to provide multiple, sometimes conflicting benefits, such as clean drinking water and food production (Nisbet et al. 2022). This is where bringing natural capital into spatial planning is important, as it can enable decision-makers to assess quantified trade-offs associated with alternative management choices, identify areas where investment in natural capital can enhance ecosystem service benefits for people (Dasgupta 2021) and better understand context dependency, that is the way in which actions vary according to wider environmental conditions (Spake et al. 2019). More research is needed into the potential of big data and associated techniques to enable spatially targeted management, and support conservation and restoration through careful land-use planning to balance economic, social and environmental trade-offs.

The condition of assets is of interest because it is likely to affect the benefits delivered, and thus the value of the asset. The condition or status of an asset can be seen as having three aspects: quantity, quality and spatial configuration (Natural England 2018, Natural Capital Committee 2014b, Mace et al. 2015). Quantity refers to the amount of an asset – its area, volume or mass. Quality refers to a range of conditions of the natural assets that affect benefits through the presence or absence of certain conditions or processes. Spatial configuration refers to the location of the asset and the spatial patterning in the landscape (Mace et al. 2015). Spatial configuration influences the delivery of benefits in various important ways, for example through habitat connectivity in the case of wildlife conservation (Isaac et al. 2018), distance from urban centres in the case of the recreational benefit of green spaces (Public Health England 2014) and the distribution of woodland in a catchment for flood protection (Mitchell et al. 2015). In the case of wildlife conservation, the quality of a habitat and its level of connectivity may be as important as the total area of habitat (Lawton et al. 2010; Isaac et al. 2018). However, the three aspects of asset condition can be highly related – for example, what appears as fragmentation when viewed at a fine resolution may be considered habitat degradation at a coarser scale. This is an area that needs more research, as it has practical implications for monitoring and understanding the relationship between asset condition and benefit delivery in more detail.

The relationship between asset condition and the flow of benefits is important for decision making and management because if the state of an assets falls below a certain level, it will affect the benefits that people receive. There is thus a target level for the benefit that society requires or desires, based on a safe level of the natural asset condition. For example, we may want to be able to harvest a certain amount of timber each year for economic reasons. The simplest case is a linear relationship between asset condition and the benefits received (Figure

3 a). For example, in the case of timber production, the value of the timber is likely to decrease linearly with the quantity of timber remaining, until all the timber has been harvested. In other cases, there will be thresholds below which the benefit rapidly decreases and the self-sustaining nature of natural capital is lost, leading to a non-linear response curve between asset condition and benefit received (Figure 3 b) (Mace et al. 2015). An example of this could be the way in which forests guard against soil erosion in a river catchment. An initial degradation in quality and selective removal of trees may have little effect on soil erosion, but beyond a certain threshold or critical level the amount of erosion increases suddenly and dramatically, with large quantities of soil washed away and reestablishment of forest or other vegetation difficult without management interventions (Pinard et al. 1995). Thus, the target level of forest cover and condition may be different when considering the benefit of protection against soil erosion to that of provision of timber, as is the effect of falling below this threshold in the response curve in each case.

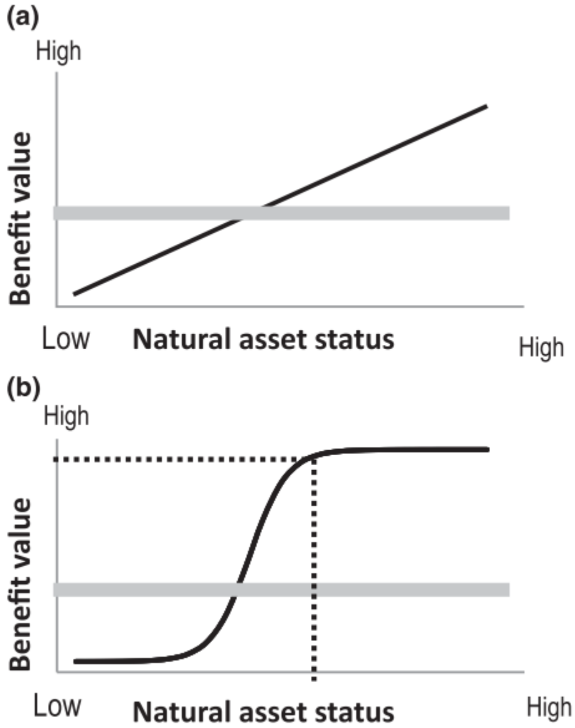


Figure 3: Different forms the asset-benefit relationship can take. The dotted line represents a threshold beyond which the decline in the benefit is very rapid. The thick grey line in each case represents a target level for the benefit, based on a safe level of the natural asset required by society. Environmental degradation can lead to a decline in the condition or status of a natural assets, moving from right to left on the x axis. In both (a) and (b) the benefit received reduces with asset condition, but in (b) the benefit received reduces very quickly after the threshold is reached, whereas in (a) the decline is slow and consistent. Taken from Mace et al. 2015.

Understanding thresholds in the response curve between asset condition and benefit delivery are important when setting target levels for benefits and natural asset condition, to ensure that benefits are secured for society, particularly when non-linear relationships are present. In some cases, these target levels are set based on critical levels that affect human health (such as the level of Mercury or Lead in water bodies), underpinned by legislation (e.g. The Water Framework Directive), and monitored to ensure compliance (European Commission 2016).

But for other assets, such as species diversity, there is no clear critical level, and there is no simple answer as to how much species diversity is enough for society. This makes setting targets challenging.

1.2.5 From theory to practice: Natural capital as a big data challenge

A proliferation of major public and private investments in data-intensive science has expanded the possibilities for scientific discovery and led to the rise of "big data", which broadly refers to the increasing volume, variety, and velocity of data streams over the past 20 years or so (Hampton et al. 2013, Chen et al. 2014). There is no strict definition of what constitutes big data, but the term is usually applied to large-volume, complex, growing datasets with multiple, autonomous sources (Wu et al. 2014). These large volumes of complex data are not readily handled by the usual data tools and practices (Hampton et al. 2013), which has led to new developments in machine learning (Jordan et al. 2015), cloud computing (Gorelick et al. 2017), data storage, and data collection capacity (Wu et al. 2014). Operationalizing natural capital is inherently a big data problem because 1) it often involves large volumes of data, either because assessments and research are carried out over broad spatial extents , long time periods or because a large number of different assets are considered, 2) a large variety of data needs to be sourced and incorporated in order to model indicators of asset status or flow of benefits, and 3) some of the datasets available are being expanded on continually, creating a "stream" of data that require new pre-processing and analytical methods (e.g. Zhu and Woodcock 2014), including close to real-time applications (Popkin 2016).

Whilst using large, varied, growing datasets from various sources undoubtably involves challenges in terms of storage and processing, it also presents opportunities to harness the information provided by novel data types and methods to monitor natural capital assets and the benefits that flow from them. The following sections will look at recent developments in available data types, methods and platforms that are relevant to monitoring within the natural capital framework.

1.2.5.1 Data

Various sources of big data can provide information on natural assets and the benefits that flow from them, including satellite remote sensing data (Pettorelli et al. 2016), camera trap data (Norouzzadeh et al. 2018), acoustic monitoring data (Oliver et al. 2018), ecological geolocated crowdsourced data (Isaac et al. 2014) and geolocated data from social media sites (Mancini et al. 2018).

Satellite remote sensing data enable large scale monitoring of land cover change, facilitated by the free and open data policy of both the USA's Landsat mission and the European Sentinel fleet (Liu et al. 2018). In the UK, satellite imagery is already the basis for the national CEH land cover maps for 1990, 2000, 2007, 2015, 2017, 2018, 2019 and 2020 which are widely used in natural capital accounting at a national (e.g. Office for National Statistics 2023a) and local level (e.g. Rouquette 2016) to quantify and map assets. Satellite remote sensing data has been identified as potentially providing other indicators of natural asset condition and flows of benefits at large scales (Ayanu et al. 2012, Andrew et al. 2014), as satellite derived indices have been shown to correlate with ground-based measures of ecological importance, and can thus be used as a proxy where ground measurements are unavailable (Morton et al. 2015). An example of this is the use of the normalized difference vegetation index (NDVI) as a proxy for above ground net primary productivity (Morton et al. 2015), which can be used to assess degradation of assets (Meneses-Tovar 2011) or predict agricultural crop yield (Panda et al. 2010).

Motion-activated cameras (or "camera traps") are becoming a mainstream tool in conservation and ecology (Rowcliffe and Carbone 2008), enabling ecologists to study population sizes and distributions, and evaluate habitat use (Norouzzadeh et al. 2018). Digital camera traps can take millions of images (Norouzzadeh et al. 2018), and advances in automated species identification in camera trap images has reduced the reliance on experts or community volunteers to extract knowledge from these images (Wearn et al. 2017, Norouzzadeh et al. 2018). Acoustic monitoring is another technique that has been developed to collect data on animals remotely (Armitage and Ober 2010). Autonomous digital recorders are capable of recording and storing enormous datasets and it is possible to detect the calls of animal species of interest from field recordings (Browning et al. 2017).

The widespread uptake of apps and social media platforms has meant that large quantities of crowdsourced data are now available. Ecological data collected using crowdsourcing techniques, such as smart phone application software and online platforms, are increasingly being used as quantitative measures of the stock and rate of change in biodiversity to assess species' risk of extinction (Isaac et al. 2014). These opportunistic biological records are relatively unstructured but vast in quantity, so can fill gaps where long-term, standardised monitoring schemes are lacking (Lin et al. 2017). Another example of geolocated crowdsourced data is photos scraped from social media platforms, such as from the photo-sharing platform Flickr. A common use for this type of data is to use the number of photos in green space as a proxy for visitation rates, and thus as an indicator of cultural benefits (Mancini et al. 2018; Wood et al. 2013). For example, Mancini et al. 2019 used pictures of wildlife posted on Flickr to quantify wildlife watching activities in Scotland, to study the overlap between recreational and conservation value of natural areas.

Whilst large and novel datasets have exciting applications within the natural capital framework, particularly for monitoring at large scales, it is important to highlight that they do not replace more traditional datasets collected in the field. Indeed, analysis using novel and large datasets are almost always combined with or compared against data collected in the field (e.g. Wood et al. 2013). In the light of current gaps in environmental monitoring and the decision to scale back the Countryside Survey (Countryside Survey 2019), it is important not to overpromise on what big data can deliver, especially in the absence of field measurements to "ground truth" results.

1.2.5.2 Methods

Harnessing the potential of big data poses challenges when it comes to pre-processing and analysis, which has led to a range of advances in methodology. Statistical techniques have been developed to extract patterns of change from noisy ecological data collected by volunteers (Isaac et al. 2014). As described previously, data collected using crowdsourcing techniques, or "opportunistic data", are increasingly being used as quantitative measures of the stock and rate of change in biodiversity to assess species' risk of extinction (Maes et al. 2015). These opportunistic biological records have an inherent sampling bias, referred to as "variation in recorder activity", due to uneven recording intensity over time and space, uneven sampling effort per visit and uneven detectability (Isaac and Pocock 2015). In order to remove this noise from the data various statistical techniques have been developed, including filtering the data to remove bias, applying a statistical correction procedure to treat recorder activity and using occupancy–detection models to estimate the conditional probability that a species is recorded when present (Isaac et al. 2014), although none of these methods corrects for all forms of variation in recorder activity.

Many studies are exploring the potential of machine learning as a method to extract information from large datasets. Machine learning is a method within the area of artificial intelligence in which algorithms learn from data, rather than being programmed manually by anticipating the desired response for all possible inputs (Jordan et al. 2015). The most widely used machine-learning methods are supervised learning methods (Hastie et al. 2017), which use predefined input-output pairs and learn how to derive outputs from inputs (Willcock et al. 2018), without being explicitly programmed. Machine learning is increasingly being used to study ecosystem services (Scowen et al. 2021), for example, Willcock and colleagues (2018) have shown that machine learning techniques can be successfully used to model ecosystem service flow. They used the Weka machine learning algorithms to model firewood use in South Africa, with comparable accuracy as conventional modelling techniques (54–77% accuracy). They also modelled biodiversity value in Sicily and highlight the benefit of the uncertainty information provided by the machine learning algorithms to help inform decision making (Willcock et al. 2018). Machine learning techniques are also a well-established way to create land cover maps using satellite data (Maxwell et al. 2018) and are central to recent developments in automated analysis of data from camera traps (Tabak et al. 2019) and acoustic monitoring (Browning et al. 2017). More recently deep learning has emerged as a "family" of machine learning algorithms, which show superior performance in some situations, and are starting to be used in ecosystem science (Perry et al. 2022), for example to better understand cultural preferences for biodiversity (Havinga et al. 2023) and map habitats using remote sensing data (Kattenborn et al. 2021). However, some challenges still remain in the use of deep learning approaches, such as the need for large quantities of reference data to train the model (Safonova et al. 2023), the technical knowledge needed to work with deep learning models, and the way deep learning models are seen as a "black box" by many, which makes the results harder to trust (Pichler and Hartig 2023).

The rapid spread of digital media and the digitization of a substantial proportion of the world's historical written resources has led to the rise of culturomics, which is the study of human culture through the analysis of changes in word frequencies in enormous digital text databases (Michel et al. 2011). Words can provide insights into human–nature interactions and through this provide new metrics and tools for near-real-time environmental monitoring and to support conservation decision making (Ladle et al. 2016). For example, Proulx et al. (2014) used Google Trends to report the seasonal trend of internet search terms, such as mosquitoes and pollen, to convey information about the biotic environment. Culturomics can also provide information on cultural benefits, by calculating the relative internet representation (a proxy of cultural saliency) of different species (Ladle et al. 2016).

1.2.5.3 Software and platforms

Uptake of the data and methods described in the previous section has been aided by an increase in open source software, including widely used programs such as QGIS, but also software that is designed specifically with the natural capital framework in mind, such as Artificial Intelligence for Ecosystem Services (ARIES 2019, Villa et al. 2014) and the other natural capital assessment tools described in Section 1.2.3.3. It is also becoming increasingly common to provide code as supplementary material in publications, which helps to disseminate new methods.

One of the challenges that emerges when working with big data is the need for a large amount of processing power and storage. This is particularly a problem for those working in the not-for-profit sector, as they usually have limited ability to pay for access to high-performance computing systems. In some areas free platforms are emerging that allow analysis to be run on a cloud. One example of this is Google Earth Engine, which is a free cloud-based platform that enables access to high-performance computing resources for processing very large geospatial datasets (Gorelick et al. 2017). There are still some concerns about data ownership and control when using Google Earth Engine, but alternatives are also appearing outside of a commercial setting. For example, the European Space Agency's Research and User Support Service currently provides free virtual machines to analyse remote sensing data (Copernicus Research and User Support 2023) as well as other web-based services that allow exploration, visualization, and analysis of datasets without having to download them (e.g. JupyterLab) (Copernicus Data Space Ecosystem 2023).

There is an abundance of "blue-sky thinking" about online platforms that leverage the new technologies associated with big data and artificial intelligence to create a "dashboard for the planet", with up-to-the-minute data on natural capital assets continuously feeding in (Green 2018). Whilst this remains for the most part a dream of those working in big tech companies, there are examples of projects that are moving in this direction, such as the open-source web application Global Forest Watch that monitors global forests in near real-time (World Resources Institute 2014). Whilst the quality of the data it provides is variable, it remains a good example of the power of platforms to harness the potential of big data whilst reaching a broad audience.

1.2.6 Research aims and thesis structure

Developing a better understanding of the link between natural capital assets and the benefits they provide society is key to support evidence-based decision making and efficiently manage our natural capital assets. To better understand this relationship and operationalise natural capital accounting at a national level it is necessary to fill in some of the gaps in our current ability to monitor assets and benefits at a national scale.

My first data chapter focuses on the flow of benefits from conservation areas in England. As described in Section 1.2.4, a full valuation of the benefits that flow from natural capital assets is important for decision making and requires indicators to help capture the different flows of benefits. Whilst the principal aim of designated areas is to protect biodiversity, the cultural flow of benefits from designated areas is not always recognised and is difficult to capture at large scales. As described in section 1.2.5, big data offers new opportunities for indicators to be developed to fill gaps in what can be monitored at a national scale. Chapter 2 will explore the potential of an emerging dataset from the collaborative encyclopaedia Wikipedia to capture and communicate the diverse socio-cultural values of nature, building on the strengths of more established crowdsourced data (Figure 4a).

	(a) Chapter 2	(b) Chapter 3	(c) Chapter 4
Operationalisation of the natural capital framework	• Full valuation of the natural world is challenging, and a range of indicators are needed • This chapter explores a new dataset to capture the sociocultural values of nature at broad spatial scales	• Society receives multiple, sometimes conflicting, benefits from natural capital assets • This chapter uses statistical methods and big data to enable spatially targeted management, to maximise benefit provision in the face of conflicting benefits	• Quantity, quality and spatial configuration of natural capital all need to be monitored to describe asset condition • This chapter explores to what extent the three dimensions of natural capital may be closely related, providing some redundancy in information
Assets and benefits considered	• Asset: designated areas in England • Benefit: Sociocultural value	• Asset: river catchments in England • Benefits: Agricultural food production and good water quality	• Asset: green spaces in Greater London • Benefits: Multiple
Data used	• Location of Flickr photographs • Geotagged Wikipedia pages • Characteristics of designated areas (coastal location, species richness, population density, etc)	• Total Oxidised Nitrogen and Reactive Phosphorous concentrations • Catchment characteristics (land cover maps, agricultural census data, population equivalent of WWTWs in the catchments, etc)	• Maps of green spaces in London • Remote sensing images • Species records • Map of tree canopies
Key aim(s)	• I compare Flickr and Wikipedia data in their ability to characterise the sociocultural value of designated areas in England • I assess how the two datasets relate to species richness.	• I assess population density as a driver of the relative importance of agricultural land use and WWTWs in predicting N and P concentrations nationally	• I assess the strength of dependency between the three dimensions of natural capital (quantity, quality and spatial configuration), using multiple indicators for each dimension

Figure 4: Overview of how the three research chapters in this thesis contribute to the operationalisation of the natural capital framework, including a brief summary of the assets and benefits they address, the data used and the key aims of each chapter.

Chapter 3 focuses on water quality in streams in England, specifically concentrations of nitrogen (N) and phosphorus (P). As identified in Section 1.2.4, the way in which assets come together to provide multiple benefits is an understudied area of natural capital. In Chapter 3, I will explore the potential of diverse data on the environment and statistical techniques to model this complexity, and to support spatially targeted management to balance the conflicting benefits of good water quality and agricultural production at a national scale in England (Figure 4b).

Chapter 4 focuses on green spaces in urban areas and examines the interdependency of the three different dimensions used to describe asset condition – quantity, quality and spatial configuration – that are widely used to assess the condition of natural capital assets. As described in Section 1.2.4, the need to monitor three different aspects for each asset significantly adds to the complexity of monitoring efforts, and yet no systematic study of their interdependency has been carried out. My analysis will use Greater London as a case study, and assess to what extent indicators of the quality, quantity and spatial configuration of green space are correlated (Figure 4c).

The thesis closes with a discussion of the implications of the work presented in this thesis, highlighting the lessons learnt, limitations of the work, technical challenges encountered and future opportunities within the study area.

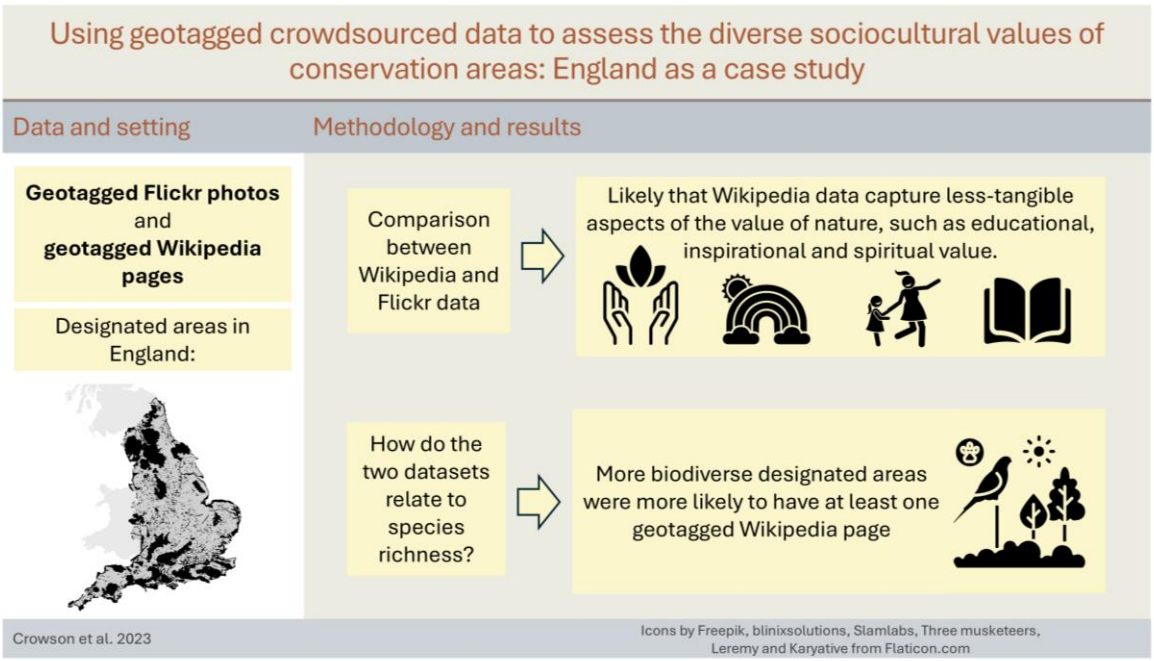
Together, I hope this work will contribute to natural capital accounting efforts by addressing knowledge gaps that hinder the operationalisation of the natural capital framework at a national scale, including ways in which new sources of data can be harnessed to provide a more complete valuation of assets, how diverse environmental data can be harnessed to manage trade-offs between different benefits through spatially targeted management, and the extent to which the three different aspects of asset condition are dependent. The progress made throughout the three chapters in understanding asset-benefit relationships will support decision making efforts to maximise the benefits we receive from nature whilst ensuring sustainable stewardship of the natural capital assets that underpin them.

Chapter 2: Using geotagged crowdsourced data to assess the diverse sociocultural values of conservation areas: England as a case study

The work presented in this chapter has been published in the following journal paper:

Crowson, M., Isaac, N. J. B., Wade, A. J., Norris, K., Freeman, R., and Pettorelli, N. (2023). Using geotagged crowdsourced data to assess the diverse socio-cultural values of conservation areas: England as a case study. Ecology and Society, 28(4):28. https://doi.org/10.5751/ ES-14330-280428

Visual Abstract



Abstract

Humanity benefits immensely from nature, including through cultural ecosystem services; geotagged crowdsourced data provide an opportunity to characterize these at large scales. Flickr data, for example, have been widely used as an indicator of recreational value, while Wikipedia data are increasingly being used as a measure of public interest, potentially capturing often overlooked and less-tangible aspects of sociocultural value (such as educational, inspirational and spiritual value). So far, few studies have explored how various geotagged crowdsourced data complement each other, or how correlated these may be, particularly at national scales. To address this knowledge gap, we here compare Flickr and Wikipedia datasets in their ability to help characterise the sociocultural value of designated areas in England, and assess how this value relates to species richness.

Our results show that there was at least one Flickr photo in 35% of all designated areas in England, and at least one Wikipedia page in 60% of them. The Wikipedia and Flickr data were shown not to be independent of each other and were significantly correlated. Species richness was positively and significantly associated with the presence of at least one geotagged Wikipedia page; more biodiverse designated areas, however, were not any more likely to have at least one Flickr photo within them. Our results highlight the potential for new, emerging datasets to capture and communicate the sociocultural value of nature, building on the strengths of more established crowdsourced data.

2.1 Introduction

Historically, arguments for conservation have promoted intrinsic values within a "nature for itself" framing (Mace 2014), but contemporary debate emphasizes the specific, quantifiable benefits that society receives from nature (Hungate and Cardinale 2017, Pan and Vira 2019). This shift in focus is illustrated by the growth of natural capital accounting both internationally (SEEA 2013) and nationally (Natural Capital Committee 2019, The White

House 2022). Natural capital is another term for the stock of renewable and non-renewable natural resources (e.g. plants, animals, air, minerals, freshwaters) that combine to yield a flow of benefits to people (Mace et al. 2015). Natural capital accounting is an umbrella term covering efforts to use an accounting framework to measure and report on natural capital and the flow of benefits we receive from it in a systematic way (SEEA 20). As an approach it emphasizes the process of valuation, namely estimating the relative importance, worth, or usefulness of natural capital to society (Natural Capital Coalition 2016). This usually involves some form of quantification, if not monetisation, to be used within decision making and planning. The goal of conducting valuations of nature is to determine in which ways nature is valuable and for whom, typically to enable better governance (TEEB, 2010, IPBES 2022). Critics of natural capital accounting highlight that the value of nature can be considered infinite and boiling this down to a series of benefits means essentially "selling out" on nature (McCauley 2006, Schröter et al. 2014). Proponents of the natural capital approach argue that if the benefits provided by nature are not assigned a value they will, by default, be assigned a value of zero, as so often happens within interactions between society and life-supporting ecosystems (Mace 2019).

A full valuation of our natural environment is challenging, however, as it underpins every aspect of human well-being, and different values emerge from different world views (IPBES 2022). A range of different metrics are needed to reflect the diverse values of nature (Harrison et al. 2017, IPBES 2022), and many of these metrics still need to be developed to operationalise the approach. Cultural ecosystem services, which include non-material benefits such as spiritual enrichment, cognitive development, recreation and aesthetic experiences (Millenium Ecosystem Assessment 2005), are usually hard to quantify and are often omitted from the valuation process despite being an important aspect of social-ecological systems. A range of methods have been developed to study cultural ecosystem services and the values associated with them, including contingent valuation (willingness to pay), choice experiments

(Cheng et al. 2019), questionnaires (Schirpke et al. 2022), deliberative methods (Allen et al. 2021), interviews, photo elicitation of values (Graves et al. 2017) and participatory mapping (Jaligot et al. 2019, Muñoz et al. 2020). These methods shed important light on people's relationship with nature, but as "stated preference" methods they rely on capturing an accurate account of people's preferences and values, which is not always straightforward (e.g. Nassauer 1983, Häfner et al. 2018). In addition, whilst these methods have important strengths at a local scale, they are very difficult to apply nationally, due to practical considerations around recruiting enough participants.

User-generated digital data has the potential to characterize diverse sociocultural values at large scales. Various studies have shown the potential of geotagged crowdsourced data from social media sites, such as Flickr, as an indicator of nature-based recreation at a national and regional scale (Wood et al. 2013, Graham and Eigenbrod 2019, Muñoz et al. 2020). This has led to a body of work on valuing cultural ecosystem services focusing on visitation rates, and the spatial and temporal variation in human engagement with the natural environment (see e.g. Van Zanten et al. 2016, Mancini et al. 2018, Calcagni et al. 2019). New opportunities are emerging to identify digital data that have the potential to characterize human perceptions of nature at large scales (Ladle et al. 2016, 2019, Schuetz and Johnston 2021) and capture public interest in species and ecosystems. A range of studies have looked at how people's interest in particular species has varied over time, using data on Wikipedia page views (Millard et al. 2021) or Google Trends (Schuetz and Johnston 2021). The collaborative encyclopaedia Wikipedia offers a powerful data source to map public interest at large spatial scales, making use of the activity of a huge community of existing users. Many Wikipedia pages are geotagged and these can be mapped to see what areas or landmarks are of interest to the public. Recent work, for example, modelled public interest in protected areas in Brazil using Wikipedia page views (Guedes-Santos et al. 2021).

However, it is not currently clear how the public interest in Wikipedia pages relates to different sociocultural values. There is some evidence that Wikipedia pages relating to attractions and events can predict visitation (Khadivi et al. 2016), but some users may decide to read and write Wikipedia pages because of an interest in a place or topic that is not directly related to visitation. This means that Wikipedia data could potentially capture a range of other non-tangible sociocultural values, such as educational, inspirational, aesthetic and spiritual value, sense of place and cultural heritage (Hernández-Morcillo et al. 2013), that are often systematically overlooked due to being 'intangible' or 'messy' benefits (Milcu et al. 2013, Chan et al. 2020).

There are currently no large-scale studies comparing the informational signature of Wikipedia data with the information contained in other geo-tagged datasets known to directly correlate with visitation rates, such as Flickr. Yet doing so would help identify what aspects of the sociocultural value of nature Wikipedia data is able to capture. To address this gap, we here compare Flickr and Wikipedia data in their ability to characterise the sociocultural value of designated areas in England. We also assess how they each relate to species richness. We chose England as there is good data on the natural environment readily available and widespread use of both Flickr and Wikipedia.

2.2 Study area

The scope of this study is limited to terrestrial ecological systems and includes designated areas on mainland England. The designation types considered are National Parks, Areas of Outstanding Natural Beauty (AONBs), Ramsar Sites, Special Areas of Conservation (SACs), Special Protection Areas (SPAs), Local Nature Reserves (LNRs), National Nature Reserves (NNRs) and Sites of Special Scientific Interest (SSSIs) (n = 6349; Figure 5). These types of designations were chosen as the most relevant for nature conservation in England following Lawton and colleagues (2010). National Parks and AONBs are designated for their cultural, landscape and (in the case of National Parks) recreational value, but also have nature conservation as part of their primary statutory purpose (Lawton et al. 2010). Ramsar Sites, SACs, SPAs, LNRs, NNRs and SSSIs have nature conservation as their primary designation purpose and have a high level of protection. Three of these types of statutory sites are a result of international treaties and obligations (Ramsar Sites, SACs, and SPAs). SSSIs and NNRs include some of the highest quality wildlife areas, whilst LNRs are designated by local authorities. There are spatial overlaps between some of the designations, for example 24% of the total area of National Parks and 12% of the area of AONBs are also designated SSSI (Lawton et al. 2010). Designated areas with the exact same extent and with more than one designation were only included once. Designated areas with offshore areas were clipped to the coastline (at low tide).

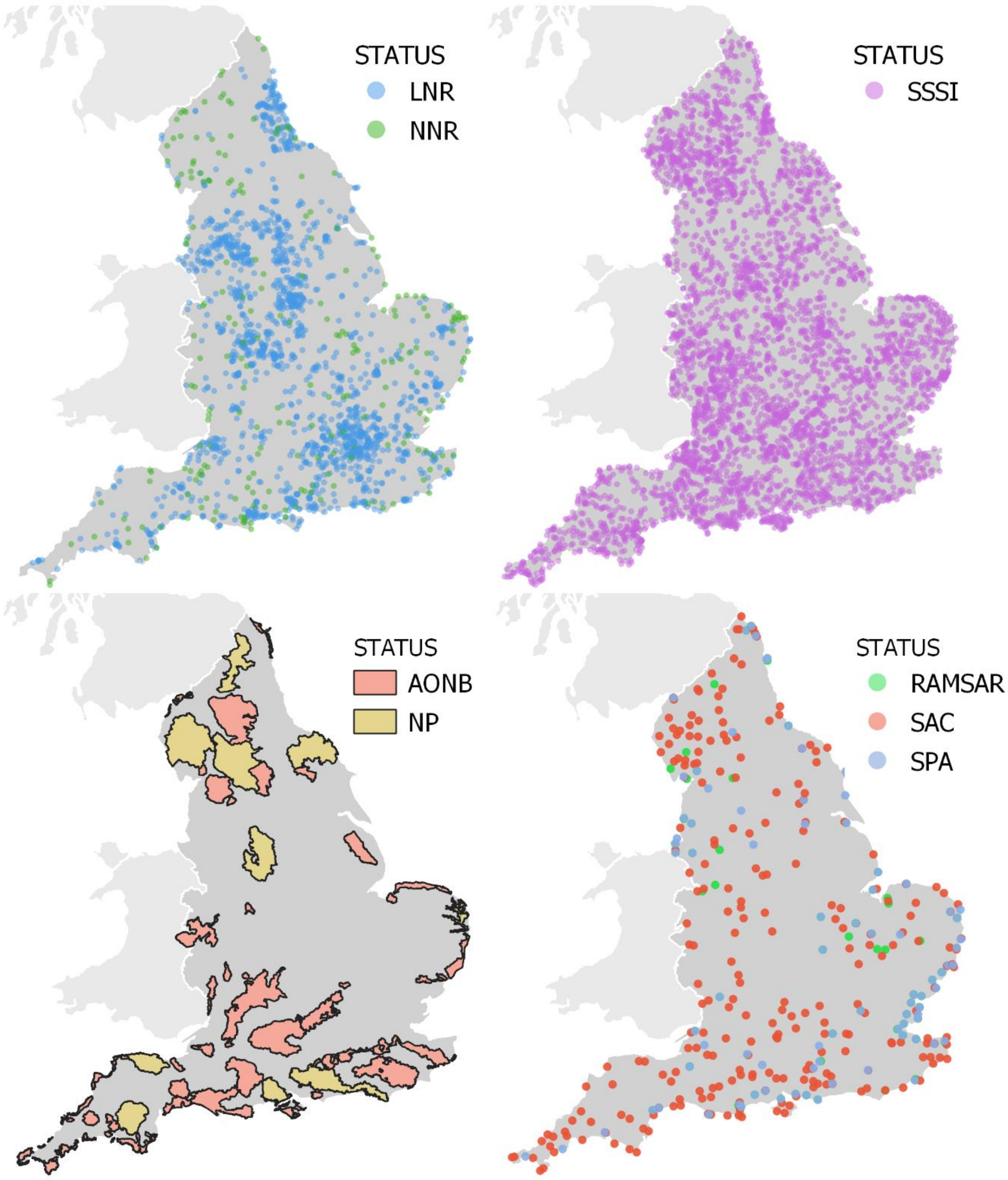


Figure 5: Map of designated areas considered in this study, broken down by designation type. Designation types considered are Local Nature Reserves (LNR), National Nature Reserves (NNR), Sites of Special Scientific Interest (SSSI), Areas of Outstanding Natural Beauty (AONB), National Parks (NP), Ramsar Sites (RAMSAR), Special Areas of Conservation (SAC) and Special Protection Areas (SPA).

2.3 Materials and methods

2.3.1 Data acquisition

To create a metric from the Flickr data, we assessed whether, or not, a given designated area had at least one Flickr photo associated with it; to create a metric from the Wikipedia data, we assessed whether, or not, a given designated area had at least one geotagged Wikipedia page within it. To assess if each designated area in England had at least one Flickr photo associated with it, the Flickr application programming interface (API) was queried to create a dataset of photos taken within designated areas between 2016 and 2019, which coincides with the period for which reliable Wikipedia data was available. The packages RCurl (Temple Lang 2020a), XML (Temple Lang 2020b) and httr (Wickham 2020) were used in R (R Core Team 2020) to request and download the data. The user and photograph ID, the date when the photo was taken and the geographic coordinates of where it was taken were downloaded. These data are anonymous, with the user ID not personally identifying the user in any way, in line with data protection regulation and data privacy concerns (Di Minin et al. 2021). Designated areas were categorised into two groups: those with no Flickr photos and those with one or more Flickr photos.

To assess if each designated area in England had at least one geotagged Wikipedia page within it, a dataset was created containing all Wikipedia pages (written in the English language) with geotags within designated areas on the 1st of May 2019. The number of Wikipedia pages has increased from year to year since at least 2004 (Wikimedia Statistics 2021), so the pages that exist on a particular day broadly represent those created in the period leading up to that date. The most recent list of unique pages and their associated geotags is available for the English-language edition of Wikipedia as a dump (https://dumps.wikimedia.org/enwiki/). A spatial filter was applied in QGIS to select the pages within designated areas. As with the Flickr photos, the areas were categorised into two groups, those with no Wikipedia pages and those with at least one Wikipedia page.

For the biodiversity indicator we used a freely available dataset on the estimated species richness of birds, butterflies and vascular plants at a 100 km^2 scale, which was compiled for the period 2000 to 2013 (Dyer and Oliver 2016), derived from species occurrence data from the Biological Records Centre. To create a single species richness indicator from the three taxonomic groups considered (birds, butterflies and vascular plants) we scaled each group, by subtracting the mean from the original data and dividing them by the standard deviation. We then selected the maximum value for species richness found within each designated area for each taxonomic group, and finally added the three values together.

In addition to the variables described above, we included variables that have been shown to influence cultural ecosystem service value in previous studies, namely distance to major towns and cities, population density, coastal location, public transport connectivity, elevation, presence of rivers and presence of waterbodies (Graham and Eigenbrod 2019, Mancini et al. 2019, Muñoz et al. 2020). To estimate the distance from designated areas to urban centres we obtained the vector boundaries for major towns and cities in England for 2015 (Office for National Statistics 2015) and calculated the minimum distance between each designated area and the closest major town or city. To quantify the connectivity of designated areas to urban centres, we used data from OpenStreetMap (2020) to obtain the number of bus stops and train stations within the "pedestrian shed" of designated areas (a 500 m buffer). A radius of 500 metres is considered in literature as a convenient pedestrian shed to capture proximity dynamics (Carpio-Pinedo 2014). We extracted the maximum population density for each designated area from the SEDAC Gridded Population of the World dataset for 2015, with a resolution of 30 arc-seconds (~1 km) (SEDAC 2018).

We used the Office for National Statistics shapefile of the extent of realm (coastline) to identify coastal designated areas (Office for National Statistics 2019). We extracted the maximum height within each designated area from the Shuttle Radar Topography Mission 90m digital elevation model data (Jarvis et al. 2008). We used data from OpenStreetMap (2020) to create a factor identifying whether each designated area has at least one river within it or not and at least one waterbody within it or not; waterbodies considered included lakes, reservoirs and wetlands.

2.3.2 Analysis

Pearson's chi-squared test was used to test for independence between the Flickr data and the Wikipedia data, and Spearman's Rank correlation was used to assess the direction and strength of correlation between these metrics.

We used binomial generalized linear models (glm) with a logit link function to model the probability of obtaining a Flickr picture or a Wikipedia page within a given designated area, implemented using the function glm from the core R stats package (R Core Team 2020). We chose glm as a method because interpretability of the model is a priority, and the coefficients are a robust way to gain insight into the relationships between the independent variables and the Flickr and Wikipedia data. Fixed covariates considered for both models were the log of the area of the designated area; species richness of birds, butterflies and vascular plants; the log of the distance to the closest major town or city; the log of the number of public transport links; population density; coastal location (categorical with two levels); maximum height; presence of at least one river (categorical with two levels); and presence of at least one waterbody (categorical with two levels). Akaike's Information Criterion (AIC) was used as the selection criteria for covariates to be included in our final best models. We used a stepwise approach, starting with a 'maximal' model including all the fixed covariates and conducting backward model selection (Zuur et al. 2009) using the function step in the stats package (R Core Team 2020).

All the covariates were standardised so that the coefficients were comparable. Model assumptions were verified by plotting residuals versus fitted values and against each covariate. We assessed the residuals for spatial autocorrelation by calculating Moran's I, creating a map of the residuals for visual inspection and by plotting a distance-based semivariogram. The percentage of variance explained by our best models was calculated using the rsq package (Zhang 2020). To evaluate the accuracy of our predictive model, we used the pROC package (Robin et al. 2011) to create a receiver-operating characteristic (ROC) curve and calculate the area under the curve (AUC) (Zou et al. 2007).

2.4 Results

With regards to the distribution of Flickr photos and geotagged Wikipedia pages, in total 2194 areas had at least one Flickr photo (34.6% of all designated areas) and 3829 areas had at least one geotagged Wikipedia page associate with them (60.3% of all designated areas).

In relation to comparing the Wikipedia and Flickr data to assess what sociocultural values Wikipedia data have the potential to capture, the Pearson's chi-squared test for independence between the Flickr and the Wikipedia data gave a test statistic of $\chi^2 = 522.64$ (df = 1, $p < 0.001$), meaning that the two datasets were not independent of each other. The Spearman correlation coefficient for the relationship between the Flickr and Wikipedia data was $r_s = 0.29$, with the two datasets shown to be significantly and positively correlated ($p < 0.001$). Designated areas with both a Flickr photo and a Wikipedia page made up 27.5% of the total, whilst 32.6% of all areas had neither a Flickr photo nor a Wikipedia page (Table 1).

Table 1: *Contingency table showing the relationship between the binary variables from the Flickr and Wikipedia data for all designated areas (n = 6349). The Flickr data is used to categorise designated areas into those with at least one Flickr photo (Y) and those with no photos (N). Likewise, Wikipedia data is used to split designated areas into those with at least one geotagged Wikipedia page (Y) and those with none (N).*

Wikipedia data		Flickr data: Flickr: Y	Flickr data: Flickr: N	
	Wiki: Y	1747	2082	3829
	Wiki: N	447	2073	2520
		2194	4155	6349

In relation to understanding the relationship between the value of designated areas and species richness, our best models for the Flickr and Wikipedia data explained 38% and 27% of the variability in the probability of a given designated area having a Flickr picture or a Wikipedia page associated with it, respectively. The AUC score for the best model of the Flickr data was 0.84, and for the best model of the Wikipedia data it was 0.80 (see Figure A1.1, in Appendix 1, for the corresponding ROC curves). These best models (Table 2 and Table 3) showed that designated areas with high values for species richness were significantly more likely to have at least one geotagged Wikipedia page associated with them (p-value < 0.001), but this was not the case for geotagged Flickr photos (p-value = 0.08) (see Appendix 2, Figure A2.1, for the prediction plots for species richness).

Table 2: *Estimated regression parameters, standard errors, z-values and p-values for the binomial glm of the Flickr data. Model* R^2 *is 0.38. The independent variables are the log of the geographical extent of the designated area, species richness, coastal location, the presence of a water body, population density in the designated area and the log of the number of public transport links.*

	Estimate	Std. error	*z* value	*p*-value
Intercept	-1.18	0.05	-25.44	< 0.001
LogArea	1.53	0.06	27.48	< 0.001
SpeciesRichness	-0.06	0.03	-1.72	0.08
Coastal : Yes	1.83	0.13	13.70	< 0.001
Waterbody : Yes	0.45	0.07	6.39	< 0.001
PopulationDensity	0.19	0.04	4.13	<0.001
LogPublicTransportLinks	0.25	0.05	5.34	<0.001

Table 3: *Estimated regression parameters, standard errors, z-values and p-values for the binomial glm of the Wikipedia data. Model* R^2 *is 0.27. The independent variables are the log of the geographical extent of the designated area, species richness, maximum height of the landscape in the designated area, coastal location, the presence of a river, population density in the designated area and the log of the number of public transport links.*

	Estimate	Std. error	*z* value	*p*-value
Intercept	0.80	0.05	17.17	< 0.001
LogArea	1.56	0.05	30.90	< 0.001
SpeciesRichness	0.38	0.03	11.13	< 0.001
MaxHeight	-0.25	0.04	-6.58	< 0.001
Coastal: Yes	-0.38	0.13	-3.02	0.002
River: Yes	-0.27	0.06	-4.09	< 0.001
PopulationDensity	0.15	0.04	3.51	< 0.001
LogPublicTransportLinks	-0.17	0.04	-3.81	< 0.001

Moran's I analyses suggested that spatial autocorrelation in the residuals of both best models remained significant, but was very small in the case of both the Flickr data (observed = 0.03, expected = -0.0002, p-value < 0.001) and the Wikipedia data (observed = 0.04, expected = -0.0002, p-value < 0.001). Analysis of the spatial autocorrelation using semi-variograms and subsampling our data showed that this amount of spatial autocorrelation did not affect our conclusions (see Appendix 3).

2.5 Discussion

Although Flickr and Wikipedia data have been used separately to study the relationship between humans and the natural world, we have shown for the first time that these two data sources are not independent of each other. Our results provide evidence that Wikipedia data captures patterns of visitation to designated areas to some extent, which makes sense given that people are likely to be interested in places that they plan to visit or have visited. However, the correlation between the Wikipedia and Flickr datasets was found to be relatively small, so it is likely that some of the signal from the Wikipedia data captures less-tangible aspects of the value of nature, such as educational, inspirational and spiritual value. The results also highlight that the diverse sociocultural values of nature are closely intertwined and hard to separate into neat categories, with visitation closely linked to public interest as captured by digital, user-generated data. This is relevant to the debate about whether designated areas should be "set aside" for nature, or managed as shared spaces between people and biodiversity (Adams et al. 2014), as a lack of access may lead to a fall in public interest in a site. Our work also shows that Wikipedia and Flickr data have different relationships with species richness, providing further support for the idea that the two datasets contain different signals, and highlighting that species richness has a significant positive effect on public interest in designated areas in England.

Interestingly, our results suggest that species richness is not generally an important driver of visitor numbers in designated areas in England, as measured by Flickr data. Previous research is mixed in this area, with some studies finding a positive relationship between designated landscapes (protected for their high biodiversity value) and the number of Flickr photos at a regional (Gliozzo et al. 2016) and national (Graham and Eigenbrod 2019) scale, whilst other found no evidence that designation increased the number of Flickr photos (Hornigold et al. 2016, Mancini et al. 2018). In our study geodiversity, such as coastal location and the presence of a water body, plays an important role (Table 2). Local population density and the number of public transport links both have small positive effects on the probability of finding a Flickr photo in a designated area, whilst distance to the nearest major town or city do not have a significant effect. These findings are surprising, as other studies have shown that connectivity and proximity to urban areas have a large positive influence on visitor numbers (e.g. van Zanten et al. 2016 for Europe, Mancini et al. 2019 for Scotland). However, these studies are from countries that, unlike England, are not densely populated, and where most designated areas are relatively inaccessible. Distance to towns was also not important in explaining visitor numbers in Vermont (Sonter et al. 2016), where conserved lands exist throughout the state and the maximum distance between any conserved land and a town is less than 100 km.

Using crowdsourced geotagged data to study the diverse values of high biodiversity areas is in the spirit of other recent work in "culturomics" that uses digital data to capture less tangible aspects of the relationship between humans and nature by, for example, capturing national park visitors' sentiment from social media text (Hausmann et al. 2020), track species awareness through time using Wikipedia page views (Millard et al. 2021) and using expressions in photographs to reflect an aesthetic judgment of natural areas (Do 2019). The challenge with this approach is that it can be very difficult to validate the digital data using independent, non-digital sources (Correia et al. 2021). The evidence that Flickr data is a good measure of visitation is strong (Wood et al. 2013, Mancini et al. 2018), but there is less evidence for what precise sociocultural values Wikipedia data can capture. Our study finds that local population density has a significant effect on the likelihood of finding a geotagged Wikipedia page within a designated area. People are often more interested in local entities and part of this is likely to be due to the ease with which they can be visited. Indeed, a study of online public interest in birds, measured by Wikipedia pageviews, found that those more commonly encountered in the wild attracted more pageviews (Mittermeier et al. 2021a). Given the lack of information on the motivation of Wikipedia users (Mittermeier et al. 2021b) it is hard to know what proportion of public interest is unrelated to visitation, and indeed it may be unproductive to compartmentalise value into such neat categories.

There are various methodological considerations that need to be considered when interpreting data sourced from Wikipedia and which point to other limitations for this study. Firstly, there are various considerations around the geography of Wikipedia data that have implications for the interpretation of our results. Because Wikipedia is organised by language rather than by country, language becomes the best proxy for country (Mittermeier et al. 2021b), which is why we chose to use the Wikipedia pages in English. This leads to two limitations. Firstly, there are English speakers across the entire world, whilst those who post georeferenced photographs on Flickr are more likely to have easy access to the area. Secondly, by choosing to consider only the English language Wikipedia pages we overlook the potential value from people originating from non-English speaking countries (or indeed whose favoured language is not English). More generally, there is the known bias in who contributes to Wikipedia and Flickr. In the case of Wikipedia, the demographic of editors is known to be predominantly white and male (Wagner et al. 2015), whilst high Flickr photograph density has been shown to correlate with high densities of well-educated white people (Li et al. 2013). Indeed, any study using data from internet users excludes those who do not use the internet. Failing to include certain sectors of the population when drawing conclusion about value is a clear limitation of this kind of "big data" approach, as the question "of value to whom?" remains an issue (Milcu et al. 2013, Ghermandi and Sinclair 2019, Wilkins et al. 2021). Whilst directly addressing these issues is either not technically possible or beyond the scope of this study, taking them into account is important when interpreting the results.

It is possible that the way Wikipedia data is used could affect what is being captured. Wikipedia data has already been used to create metrics and indicators in a range of ways, and our approach of using geotagged pages within designated areas is cutting edge, capturing a broad range of spatial entities within designated areas that contribute to their value, from streams to stone circles. However, future work could develop indicators that make use of additional variables associated with the online encyclopaedia, such as page size (Wong and Rosindell 2021), page name (Chua et al. 2021), page views (Nolan et al. 2022), page edits and the distribution of languages and users (Mittermeier et al. 2021a).

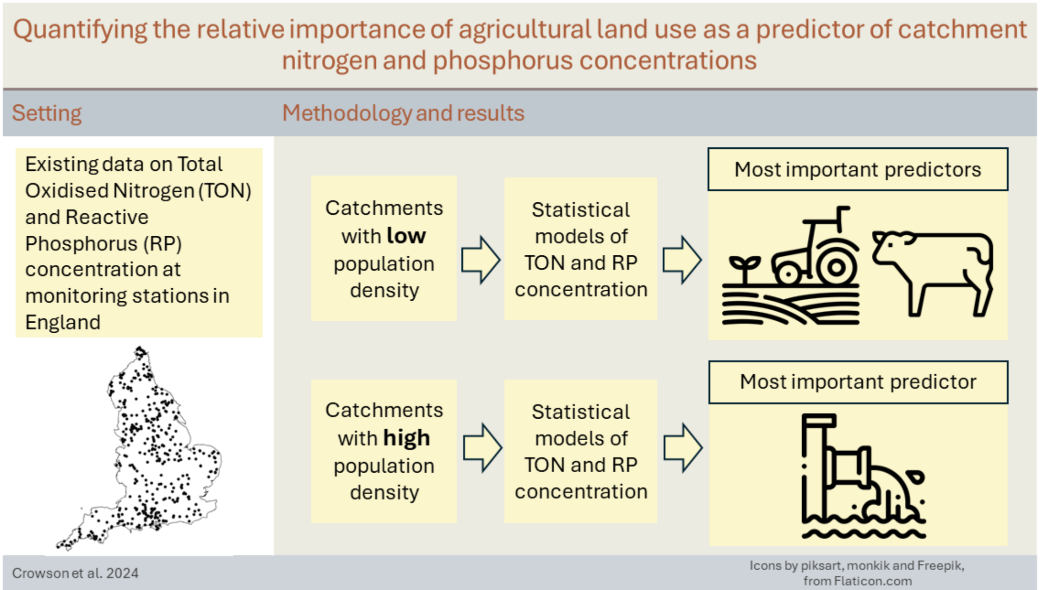
A further limitation of our study is the species richness data used to model value. The spatial resolution of this data is 100 km^2 (Dyer and Oliver 2016a), which is a course resolution given the small size of many designated areas in England (the median size of LNRs in this study is 0.1 km^2, for example). This means that variation in species richness at a scale relevant to the smaller designated areas may not be captured by the dataset, in cases where these designated areas are small hotspots of biodiversity. In addition, the taxonomic groups included (birds, bees, and vascular plants) do not include other groups that are likely to be of interest to people, for example mammals. Whilst the data used is currently the best available at a national scale, there is room for improvement should other data be published in the future. In addition, future work could explore to what extent particular "charismatic species" increase visitor numbers, rather than more general measures of species richness. There is, for example, evidence that people will pay more or stay longer in a protected area if they have the possibility of encountering particular wildlife species (Mustika et al. 2020). It is also possible that the abundance of species influences recreational value, as visitors are more likely to be able to see species that are present in large numbers. Exploring further what aspects of biodiversity people care about and are more likely to visit is a promising direction for future work in this area.

Understanding the values of nature is a fundamental step to comprehend and manage the interlinkages between people and other-than-human nature (Díaz et al. 2015). This study, as is the case for all attempts to value nature, emerges from a particular regional context and world view. The natural capital approach in England pushes for a national scale assessment of the value of stocks and the benefits that flow from them (Natural Capital Committee 2020), and regular reporting of these within national accounts by the government (Office for National Statistics 2022c). It is within this type of valuation exercise that it is important to find datasets that are relevant to national policy and which can give a static "snap shot" of the state of the environment and the values we derive from nature. However, it is possible that any national assessment of value will remain very limited, as people's experience of the natural world is primarily local. There are other ways of setting up valuation exercises that these data could also be used for, at various scales, including valuation exercises that highlight the importance of directing ecosystem service assessment and valuation exercises towards specific trade-offs and decision making (Posner et al. 2016, Chan et al. 2020). In the case of characterizing the diverse values of nature to inform environmental decision-making, the "culturonomics" approach taken within this study is probably best combined with more discursive and deliberative methodologies, both to ensure public participation and reflection (Allen et al. 2021) and to help define the cultural ecosystem services through the lens of the beneficiaries themselves (Katz-Gerro and Orenstein 2015). Finally, the fact that the Flickr and Wikipedia data used in this study have a spatial dimension means that they can be used in mapping exercises, and through this enhance the visibility and communication of cultural ecosystem services, which is another important aspect of valuation exercises (Hernández-Morcillo et al. 2013).

The intrinsic value of high biodiversity areas remains a strong moral argument for their continued conservation, however the focus within policy on including the benefits humans receive from nature into decision-making in a formal way requires new approaches to capturing value, including attempts at quantification. This study has shown that there is potential for new, emerging datasets to capture and communicate the sociocultural value of nature, building on the strengths of more established crowdsourced data. However, in addition to critically assessing new datasets as we have done in this study, it is important to recognise that no single indicator can be expected to represent the value of a landscape and can, in fact, be used against conservation efforts. For example, many of the designated areas did not have any Flickr or Wikipedia data associated with them, but this does not mean that they do not have any value to people. Going forward it would be interesting to use a series of case studies on different designated areas, at various scales, with different characteristics and designation types, and combine "stated preference" and more deliberative methods with digital data to understand in more detail the processes at work, including people's motivations.

Chapter 3: Quantifying the relative importance of agricultural land use as a predictor of catchment nitrogen and phosphorus concentrations

Visual Abstract



Highlights

- The relative importance of N and P sources in rivers was assessed nationally.
- A data-driven, statistical, approach focused on population density was used.
- Agricultural sources of N and P dominate in catchments with low population density.
- Waste water treatment works dominate in catchments with high population density.
- The findings inform spatially targeted management.

Abstract

Human society receives the benefits of food production and good water quality from its endowment of natural capital, however there can be conflict between the provision of these two benefits. Agriculture is a major source of nitrogen (N) and phosphorus (P) in freshwater ecosystems, and different management strategies exist to reduce farmland nutrient losses and thus mitigate freshwater eutrophication. The importance of agricultural sources of N and P as drivers of water quality is known to vary spatially, but quantification of the *relative* importance of the nutrient sources and transport pathways shaping this variability remains challenging, especially with reference to inputs from waste water treatment works. Addressing this knowledge gap is key for targeting management strategies to where they are likely to have the greatest effect. To advance our understanding in this area, this study assesses the impact of population density as a driver of the relative importance of agricultural land use for predicting mean Total Oxidised Nitrogen (TON) and Reactive Phosphorus (RP) concentrations in rivers in England, using two different data-driven, statistical approaches: a generalised linear model and random forest. Our results show that agricultural N and P sources dominate in catchments with low population density, where stream water concentrations are lower and waste water treatment works are numerous, but smaller in terms of the population equivalent served. Agricultural N and P sources are not important predictors of N and P in catchments with high population density, where point source contributions from waste water treatment works dominate and the mean TON and RP concentrations are higher. These results require cautious interpretation, as model validation outcomes show that high TON and RP concentrations are consistently underpredicted by both statistical approaches. Altogether, our results lend support to the idea that the relative contribution of agricultural sources may be overestimated in densely populated catchments, relative to point sources from waste water treatment works, and that management strategies to reduce the contribution of agriculture to

N and P in rivers may be better targeted towards catchments with lower population density, as this is where agricultural land use is the primary source of N and P in rivers.

3.1 Introduction

Natural capital needs to be managed in a way that balances the different benefits it provides society. Here, natural capital is another term for the renewable and non-renewable natural resources (e.g. plants, animals, air, minerals, freshwaters) that combine to yield a flow of benefits to people (Mace et al. 2015). However, these benefits are sometimes conflicting, for example, the benefit of agricultural food production can conflict with good water quality, as agriculture disrupts nitrogen (N) and phosphorus (P) cycling through organic and inorganic N and P fertiliser application, and livestock rearing, increasing stream water N and P concentrations (Howden et al. 2010). In freshwater ecosystems, increased levels of N and P can result in eutrophication, with its associated shift to plant communities dominated by fast-growing competitive species (Mainstone and Parr 2002, O'Hare et al. 2018), excess growth of aquatic weeds and phytoplankton, blooms of harmful algae and the associated negative impacts on invertebrates and fish (Smith and Schindler 2009). This in turn adversely impact on a range of water uses and societal benefits, including drinking water abstraction and treatment, livestock watering, water sports, angling, amenity value and tourism (Environment Agency 2019).

Agriculture is known to be a major source of N and P, and nutrient runoff from agricultural practice is an underlying cause of eutrophication in many catchments (Carpenter et al. 2011, Moss 2008), N and P reach streams through wash-off and leaching of nutrients from fertiliser and manure applications to arable landscapes, and through soil disturbance and sediment runoff due to land management practices and livestock grazing (Nisbet et al. 2022). A range of measures have been developed to reduce diffuse pollution from agriculture, including reduced fertiliser usage, reduced tillage, and crop rotation (Luna Juncal et al. 2023). There are concerns that these measures do not go far enough to reach water quality targets, which has led to the focus being increasingly put on land cover change, usually from crop to forest, peatland or wetland, that is, from land use that inputs N and P into river systems to one that can capture N and P (Nisbet et al. 2022). This type of natural habitat restoration often targets key areas around the sources, pathways, or receptors of N and P, whilst delivering many other benefits such as habitat creation, shade creation, carbon sequestration and increase access for recreation (Langhans et al. 2022).

However, most of the management strategies aimed at reducing diffuse agricultural sources of N and P are costly to implement and have implications in terms of reduced yield or added management effort for farmers. In some cases, farmers receive renumeration for carrying out management strategies on their land, through payment schemes funded in various ways, for example through taxes (e.g. the Environmental Land Management schemes in England) or water utility companies (Nisbet et al. 2022). Land cover change comes at both an economic and social cost, as taking agricultural areas out of production has implications for food security. To address human society's need for both food production and good water quality, it is important for management strategies to be spatially targeted to the sites where the measures will have the biggest positive effect on improving water quality in rivers (Withers et al. 2014), through an understanding of the effect of key drivers in different contexts (Spake et al. 2019).

There is some evidence that in densely populated regions the contribution of agriculture to P concentrations in rivers may be less important than previously thought (Withers et al. 2014). A comparison of 10 countries in northwest Europe showed that mean P concentration in rivers were more strongly correlated with discharges associated with urban populations than with agricultural variables (Foy 2007). In addition to this, a regional study of N concentration in an urban-dominated region showed that urban is the land characteristic which is most important in determining nitrate concentrations (Davies and Neal 2004), but when the analysis was applied to landscapes across the UK, the area of arable land proved to be more important

(Davies and Neal 2007). This could mean that catchments with a low population density are a better choice for mitigation measures targeting agricultural sources of N and P, if they are shown to be the dominant cause of nutrient enrichment in these rivers, and thus more likely to respond to interventions with improved water quality. However, it is also possible that point sources from waste water treatment works (WWTWs) dominate in catchments with low population density, but with lower concentrations of N and P than catchments with higher population density. To date, no studies have compared catchments with low population density to catchments with high population density explicitly and at a national scale, with respect to the relative contribution of agricultural sources to N and P concentration. This study aims to fill this gap by using statistical models to test a series of hypotheses, using England as a case study. Data-driven, statistical, approaches provide an interesting and useful contrast to other models that define the relative inputs or flux transfers from different nutrient sources at the outset, for example, export co-efficient modelling and similar (Johnes et al. 1996), since the statistical models determine the relationship between source and instream concentration through model fitting. England was chosen because of the availability of water quality and environmental data, and because catchments with a range of different population densities are available, including catchments with very high population densities. Based on previous work, we expected agricultural sources to be the most important predictor of N and P concentrations in catchments with low population density (Foy 2007, Davies and Neal 2007) (H1). We expected effluent from WWTWs to be the most important predictor of N and P concentrations in catchments with high population density, with agricultural sources being less important (Davies and Neal 2004, Davies and Neal 2007) (H2).

3.2 Materials and Methods

3.2.1 Data

3.2.1.1 Dependent variables

We used data from the Water Quality Data Archive (Environment Agency 2021) on concentrations of Total Oxidised Nitrogen (TON) (Total Oxidised as N in mg/l, determinand notation 116) and Reactive Phosphorus (RP) (Reactive Phosphorus as P in mg/l, determinand notation 180, "Orthophosphate"), filtered for measurements taken on a river or running surface water, and taken for monitoring purposes (as opposed to compliance). We chose these forms of N and P because they are much more commonly measured that total N and total P for monitoring purposes in England. For example, the 2019 dataset has 32,753 records for the determinand TON compared to 6,064 for Total N, and 30,875 records for RP and none for Total P. We downloaded the data for the years 2015 to 2019 and filtered all available monitoring stations within England to those that had at least one measurement per season per year for this time period, providing us with 528 monitoring stations for TON and 507 for RP. We did this as there is likely to be substantial seasonal variation in the TON and RP concentrations (Shen et al. 2020), and we wanted to make sure that this is captured within the data for all monitoring stations included in the study. We then took the mean value for all the TON and RP concentration measurements for each monitoring station across the five years. We chose to use the mean value rather than the median, as the mean concentration of N and P is currently used in relation to standards for N and P in rivers in England within policy documents (e.g. Defra 2014). We chose the period 2015-2019 after initial investigations showed that extending this period meant a drop in monitoring stations that met the criteria of having at least one measurement per season, particularly as during the COVID pandemic the number of measurements taken at some monitoring stations dropped considerably, leaving seasonal gaps.

3.2.1.2 Catchments

To create catchments for the monitoring stations included in the study we snapped the geolocation of each monitoring station to the Centre for Ecology and Hydrology (CEH) 1:50,000 Watercourse Network dataset (Moore et al. 1994) using the r.stream.snap function (Jasiewicz 2021) in GRASS GIS (GRASS Development Team 2022) with 2 km as the maximum distance tolerance. We then used the Watershed tool (Spatial Analysis) in ArcGIS Pro (ESRI 2022) in batch mode, with the Integrated Hydrological Digital Terrain Model (IHDTM) Outflow Direction raster in its native 50 m resolution (Morris and Flavin 1990, 1994) to automatically delineate a catchment for each monitoring station. The CEH Watercourse Network dataset is consistent with the IHDTM Cumulative Catchment area, so the step of snapping the monitoring station avoids spatial discrepancies between the monitoring stations and the IHDTM Cumulative Catchment Area that would lead to large mistakes in the catchment delineation step.

There were some instances in which the above process did not work, particularly in flat regions such as East Anglia. These cases were usually easy to spot as the resulting catchments were very small (< 0.05 Km^2). In these cases, the catchments were created manually through visual inspection of the data and existing maps of catchment available through The National River Flow Archive (2023) and the Defra Catchment Explorer (2023).

There were a few instances in which catchments could not be reliably defined using the methods described above, and these were removed from the dataset. In addition, a few catchments were removed because they fall mostly in Scotland and Wales, which are beyond the scope of this study and in some cases have differences in data availability. Finally, one catchment was removed because the monitoring station was immediately downstream from a fertiliser factory and had extremely high values for TON concentration. This process left a total of 515 monitoring stations to model concentrations of TON and 494 monitoring stations to model concentrations of RP. However, many of these catchments overlap, that is, they contain each other due to them being on the same river or branch of a river. The observations at monitoring stations that are downstream from each other are not independent from each other (Schreiber et al. 2022), as the water passing through them will be affected by the same conditions, processes and events, leading to pseudoreplication, which is an issue when interpreting the models used in this study (Mets et al. 2017). To avoid this bias, we grouped the catchments that overlap, and selected the catchments with the highest elevation outlet within each group. This means that there is a bias towards catchments with a greater ratio of upland to lowland land cover types, but it maximises the number of non-overlapping catchments. This is because, for example, it was possible to keep various monitoring stations with catchments on different branches of a river by removing a monitoring station lower in the landscape that contained the above. This process left a total of 404 monitoring stations to model concentrations of TON and 383 monitoring stations to model concentrations of RP (Appendix A4, Figure A4.1). As can be seen in Figure A4.1 in Appendix 4, most of the monitoring stations are included in both datasets (379), with a few only included in the TON dataset (25) or the RP dataset (4).

3.2.1.3 Independent variables

We chose the independent variables to include in the models for TON and RP based on environmental characteristics of the catchment that may impact on the concentration of N and P in rivers, namely proportion of the catchment with arable and horticultural land cover, proportion of area covered by forest, mean population density in the catchment, cattle and calf density in the catchment, sheep and lamb density in the catchment, maximum mean precipitation in the catchment, mean slope in the catchment, channel density for the catchment, the estimate of the base flow index based on the Hydrology of Soil Types classification (BFIHOST), proportion of the catchment designated for conservation and/or recreation, catchment area and population equivalent of the WWTWs within the catchment. Population equivalent is a parameter for characterizing the polluting potential of industrial wastewaters (in terms of biodegradable organic matter). For the models of TON, we also included the mean atmospheric deposition of N. For the model of RP, we assumed atmospheric P deposition is relatively low, occurring only from windblown dust, and is unlikely to show any systematic spatial variation at a small scale (Tipping et al. 2014). A summary of independent variables included can be found in Table 4.

Table 4: *Independent variables included in the models for TON and RP. All variables are continuous. The independent variable marked with * was only used for the models of TON.*

Variable Type	Independent variable	Abbreviation
Land cover	Proportion arable and horticultural land cover	ArableHortProp
Land cover	Proportion broadleaved and coniferous woodland land cover	ForestProp
Waste Water Treatment Works	Population equivalent of waste water treatment works in the catchment. Population equivalent is a parameter for characterizing the polluting potential of industrial wastewaters (in terms of biodegradable organic matter). It expresses the polluting load of a WWTW in terms of the population (number of people) that could produce the same polluting load.	PopEquiWWTW
Catchment size	Catchment area	CatchmentArea
Soil and geology	Estimate of the base flow index based on the Hydrology of Soil Types classification (BFIHOST)	HOSTBaseFlowIndex
Precipitation	Maximum mean annual precipitation for 2015-2019	MaxPrecipitation
Population	Population density	PopDensity
Slope	Mean slope in the catchment	MeanSlope
Atmospheric deposition*	Mean atmospheric deposition of N 2015-17*	AtmosDeposition
Channel density	Channel density	ChannelDensity
Land use	Cattle density	CattleDensity
Land use	Sheep density	SheepDensity
Land use	Proportion of catchment designated for conservation or recreation	DesignatedAreaProp

All of the data preparation steps were carried out in R (R Core Team 2022), unless otherwise stated. To calculate the proportion of each catchment with arable and horticultural land cover we used the CEH Land Cover Map of Great Britain for 2017 at 25 m resolution (Morton et al.

2020). We used the same dataset to calculate the area covered by forest in each catchment, considering the classes “Broadleaved woodland” and “Coniferous Woodland” together. We acquired a list of all WWTWs from the Environment Agency, which included the population equivalent for larger works covered by the Urban Waste Water Treatment Directive, specifically those works that serve population equivalents greater than 2000 (Environment Agency 2023). For the smaller WWTWs (those that serve population equivalents < 2000) we assigned a value of 1000 for the population equivalent, to represent the central tendency of this group, in the absence of more specific data. We mapped the WWTWs based on the grid reference of the outlet, and added together the population equivalent of all WWTWs that fall within each catchment.

Mean population density within each catchment was determined using the Output Areas from the 2011 Census for Population Density (Office for National Statistics 2011). We calculated the mean of the Output Areas within the catchment, weighted by the area of each intersection between the Output Areas and the catchment. To estimate mean cattle density and mean sheep density within each catchment we used data from the England Agricultural Census 2016 at 5 km resolution on the total number of cattle and calves, and the total number of sheep and lambs (England Agricultural Census 2016). In each case, the total number was added across the catchment and divided by the area of the Agricultural Census grids that intersect with the catchment to estimate stocking densities.

To calculate the maximum mean annual precipitation for each catchment we used the HadUK-Grid rainfall data, averaged by year, on a 1 km grid over the UK (Met Office 2020). We took the mean by grid for the years 2015 to 2019 and chose the maximum value that fell within each catchment.

Mean slope was computed for each catchment using the Slope tool (Spatial Analysis) in ArcGIS and the IHDTM Digital Elevation Model (Morris and Flavin 1990, 1994). To determine channel density, we used the CEH 1:50,000 Watercourse Network dataset (Moore et al. 1994) to calculate the length of channels within each catchment and divided this by the catchment's total area (Rahman and Rahman 2020). To account for the soil and geology we calculated a base flow index (BFIHOST) for each catchment. We used the Hydrology of Soil Types (HOST) dataset (Boorman et al. 1995, Griffin et al. 2019) as the basis for our calculations, following the area-weighting method in the Flood Estimation Handbook volume 5 (Bayliss 1999, Griffin et al. 2019).

To calculate the proportion of each catchment designated for conservation or recreation, we acquired the shapefiles for terrestrial designated areas on mainland England, based on those described in Lawton et al. (2010). The designation types considered are National Parks, Areas of Outstanding Natural Beauty, Ramsar Sites, Special Areas of Conservation, Special Protection Areas, Local Nature Reserves, National Nature Reserves and Sites of Special Scientific Interest (SSSIs) (n = 6349).

For the models of TON concentration, we included the mean atmospheric deposition of N in the catchment, based on N deposition data at 1 km resolution, from the UK CEH Environmental Information Data Centre (Tomlinson et al. 2020). We used the period 2015-17, as this data was not available for after 2017. We took the mean value from all points within the catchment for each of the four forms of atmospheric deposition (dry deposition of reduced N, dry deposition of oxidised N, wet deposition of reduced N and wet deposition of oxidised N) and added them together to produce a single value.

3.2.2 Analysis

Our analysis uses catchments characteristics to explain the variation in N and P concentrations at monitoring stations at a national scale. The dependent variable used in the statistical models is either the mean TON or mean RP concentration at monitoring stations between 2015-2019 (as described previously in Section 3.2.1.1). Thus, each row in the dataset represents a monitoring station, and there is a single summary value of TON and/or a single summary value of RP for that site. The independent variables are a series of catchment characteristics summarised to a single value for each monitoring station's catchment (such as proportion of different land cover types, mean slope, etc, as described previously in Section 3.2.1.3). Agricultural land use is represented by three different variables: proportion of catchment with arable and horticultural land cover, cattle density and sheep density, representing the main agricultural sources of N and P in England (Defra 2024a and b).

To test our first hypothesis, we selected the catchments from the TON dataset with a population density below the first quantile for population density (population density < 0.41 people/ha, n = 101, Figure 6) to create a group of catchments to represent low population conditions. We did the same thing for the RP dataset (population density < 0.40 people/ha, n = 96, Figure 6).

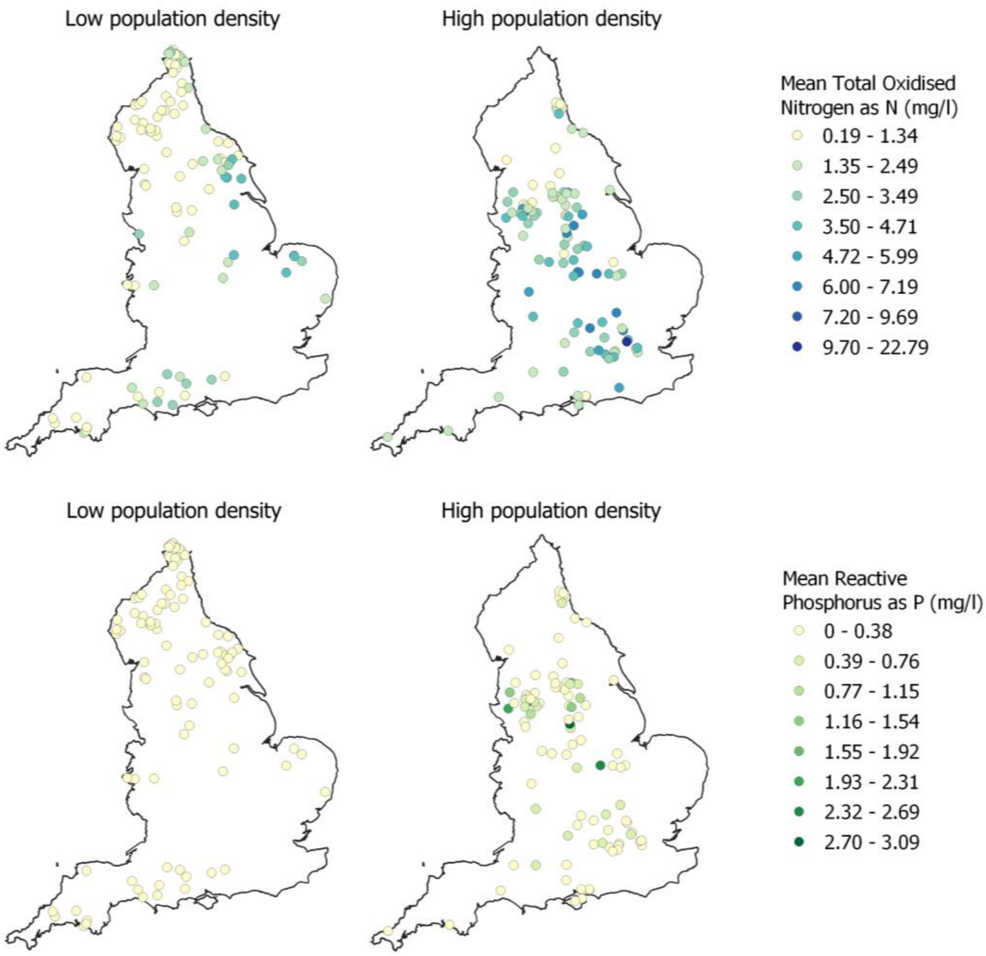


Figure 6: Mean TON (top two maps) at monitoring stations with catchments with low population density and catchments with high population density for the years 2015-2019 (n = 101 in each case). Mean RP (bottom two maps) at monitoring stations with catchments with low population density and catchments with high population density for the years 2015-2019 (n= 96 in each case). Catchments with low population density are those below the first quantile for population density in the full dataset (population density < 0.41 people/ha for TON, population density < 0.40 people/ha for RP), and catchments with a high population density are those above the fourth quantile for population density in the full dataset (population density > 3.61 people/ha for TON, population density > 3.25 people/ha for RP).

To test our second hypothesis, we selected catchments from the TON dataset with population density above the fourth quantile for population density (population density > 3.61 people/ha, n = 101, Figure 6) to create a group of catchments to represent high population conditions. We repeated the process with the RP dataset (population density > 3.25 people/ha, n = 96, Figure 6). We chose the lower and upper quantile as cut off points for our two groups of catchments because this allows us to look for differences between two strongly contrasting groups.

Histograms of the distribution of the dependent variable for each of these four datasets can be seen in Appendix 4, Figure A4.2. We ensured that the independent variables were not strongly correlated in each case (Pearsons correlation coefficient < 0.75), as this is a requirement when interpreting the statistical methods used in this paper. The distribution of the independent variables for each of the four datasets can be found in Appendix 4 (Figure A4.3 for TON and Figure A4.4 for RP).

We used two methods to model each of the four datasets: 1) negative binomial generalised linear model and 2) random forest. We chose to use both a negative binomial generalised linear model and random forest model in each case to check for consistency of results across models, and to make use of the different strengths of the two approaches. Generalised linear models are a common tool in environmental science, and are highly interpretable, as the coefficients are a robust way to gain insight into the relationships between the independent and dependent variables, and the relative importance of the dependent variables (Zuur et al. 2009). Random forest is a widely used machine learning technique, including in environmental science (e.g. Cutler et al. 2007, Molnar 2023, Ross et al. 2021), largely due to its ability to deal with nonlinear interactions and excellent predictive capability (Yu et al. 2021). Random forest has been successfully used in the past to model and predict N and P concentrations at a national scale in the USA (Shen et al. 2020).

3.2.2.1 Model training and validation

We split each dataset into training and test data using a 70/30 split, and used the same training and test dataset for both the generalised linear model and random forest model in each case. We used the negative binomial model with a log link function to model the concentrations of RP and TON, implemented using the function glm.nb in the MASS package in R (Venables and Ripley 2002, R Core Team 2022). We converted the concentrations to integer values (multiplication by 1000). Fixed covariates considered for the models of TON and RP were the proportion of arable and horticulture land cover within the catchment; log of the proportion of forest land cover; log of the population equivalent of the waste water treatment works within the catchment; population density; log of the density of cattle in the catchment; log of the density of sheep; log of the maximum mean precipitation; the mean slope in the catchment; channel density; the HOST base flow index; log of the proportion of the catchment that is designated for conservation and/or recreation; log of the total area of the catchment. The atmospheric deposition of N was also included as a fixed covariate for the model of TON.

Akaike's Information Criterion (AIC) was used as the selection criteria for independent variables to be included in our final best models. We used a stepwise approach, starting with a 'maximal' model including all the fixed covariates and conducting backward model selection (Zuur et al. 2009) using the function stepAIC in the MASS package (Venables and Ripley 2002).

All the covariates were standardised (by subtracting the mean from the original data and dividing them by the standard deviation) so that the coefficients were comparable. Model assumptions were verified by plotting residuals versus fitted values and against each covariate. To build random forest models for RP and TON, we trained the randomForest function in the R package randomForest (Liaw and Wiener 2002). We set the parameter ntree to the default of 500, based on various trial runs and recommendations in the literature (Belgiu&

Drăgu 2016). We used the function tuneRF, also in the randomForest package (Liaw and Wiener 2002), to set the value for mtry (mtry = 3 for the model of TON for catchments with low population density, mtry = 4 for the model of TON for catchments with low population density, and mtry = 3 for the model of RP for catchments with low population density, mtry = 1 for catchments with high population density). We trained the random forest using the same independent variables as were selected in the best negative binomial generalised linear model, however we did not log any of the variable or scale them, as random forest is invariant to such transformations of the independent variables, and they make model interpretation more difficult.

For all eight models we calculated the root mean square error (RMSE), plotted the test data against the concentrations predicted by the model, and calculated the strength of the correlation between these observed and predicted concentrations using Pearson's correlation coefficient. The variation explained by our generalised linear models and random forest models was calculated using the test data, following the method implemented in the randomForest package, using the formula: 1 – mse / Var(y).

We assessed the model's residuals for spatial autocorrelation by creating a map of the residuals for visual inspection, calculating Moran's I, and by plotting a distance-based semivariogram.

3.2.2.2 Effect size and variable importance

For the negative binomial generalised linear model, we created effect plots for the independent variables, with all other variables kept at their mean. For the random forest models, we extracted variable importance measures using the importance function in the randomForest package (Liaw and Wiener 2002). There is little consensus in the machine learning literature on how to best calculate the relative importance of different independent variables (Yu et al. 2021), so we report two widely used methods to rank predictor variables associated with random forest: mean decrease in accuracy and mean decrease in node impurity. Mean decrease in accuracy is computed by permuting each independent variable in the random forest, comparing the prediction error using the out of bag data, and assessing the increase in error (mean square error) when each target variable is randomized (permuted) (Liaw and Wiener 2002; Yu et al. 2021). Mean decrease in node impurity is the total decrease in node impurities (residual sum of squares) from splitting on the variable, averaged over all trees. Both are implemented within the randomForest package (Liaw and Wiener 2002).

3.3 Results

3.3.1 Model validation

The Pearson correlations between predicted and observed values for the models of TON are in the range of 0.6 - 0.9 ($df = 28$, $p < 0.001$ in all cases) across the testing sets (Figure 7), and for the RP datasets they are in the range of 0.39 – 0.84 ($df = 28$, $p < 0.001$ in all cases) across the testing sets (Figure 8). The models for RP and TON underestimate the higher values in the dataset. The RMSE for the models are shown in Figure 7 and Figure 8.

The best negative binomial models for TON explain 41% of the variation in the test data for catchments with low population density and 35% in catchments with high population density. The best negative binomial models for RP explain 47% of the variation in the test data in catchments with low population density and 61% for the catchments with high population density. The variation in the test data explained by random forest is 77% for the TON model of catchments with low population density, 44% for the TON model of catchments with high population density, 5% in the case of the RP model for catchments with low population density and 27% for the RP model of catchments with high population density. Moran's I analyses on the residuals of the generalised linear model and random forest residuals shows no significant spatial autocorrelation in the residuals of any of the models (p-value > 0.05) relevant to the scale of analysis.

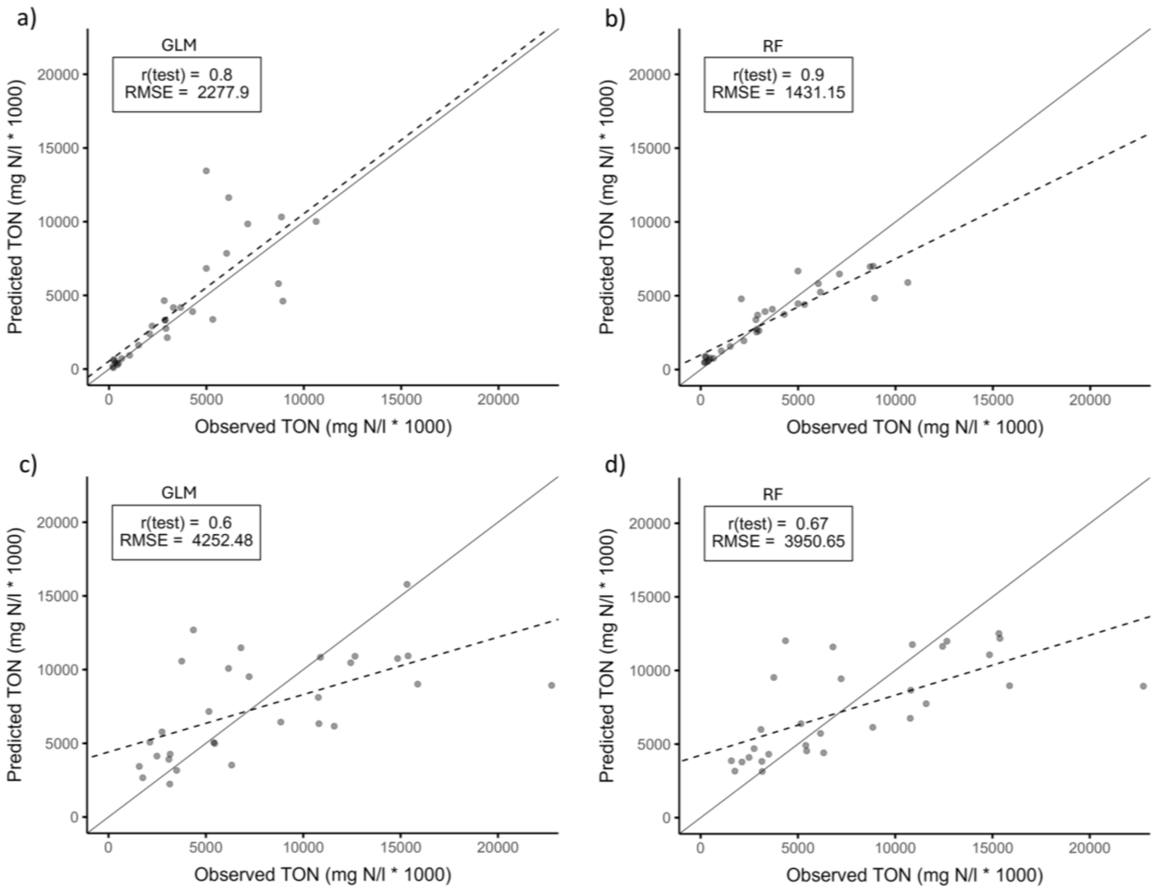


Figure 7: Correlation plots for the out-of-bag test data (n=30) for the negative binomial generalised linear model (a) and random forest model (b) of TON in catchments with low population density, and the negative binomial generalised linear model (c) and random forest model (d) of TON in catchments with high population density. Horizontal axes show the true values from the test data set, multiplied by 1000 to obtain an integer, whilst the vertical axes show the values predicted by the model. The dashed line shows the linear regression of the data points and the solid line represents the 1:1 relationship. The box in the upper left corner gives the Pearson coefficient and the value for RMSE for each model. The labels GLM and RF refer to generalised linear model and random forest respectively.

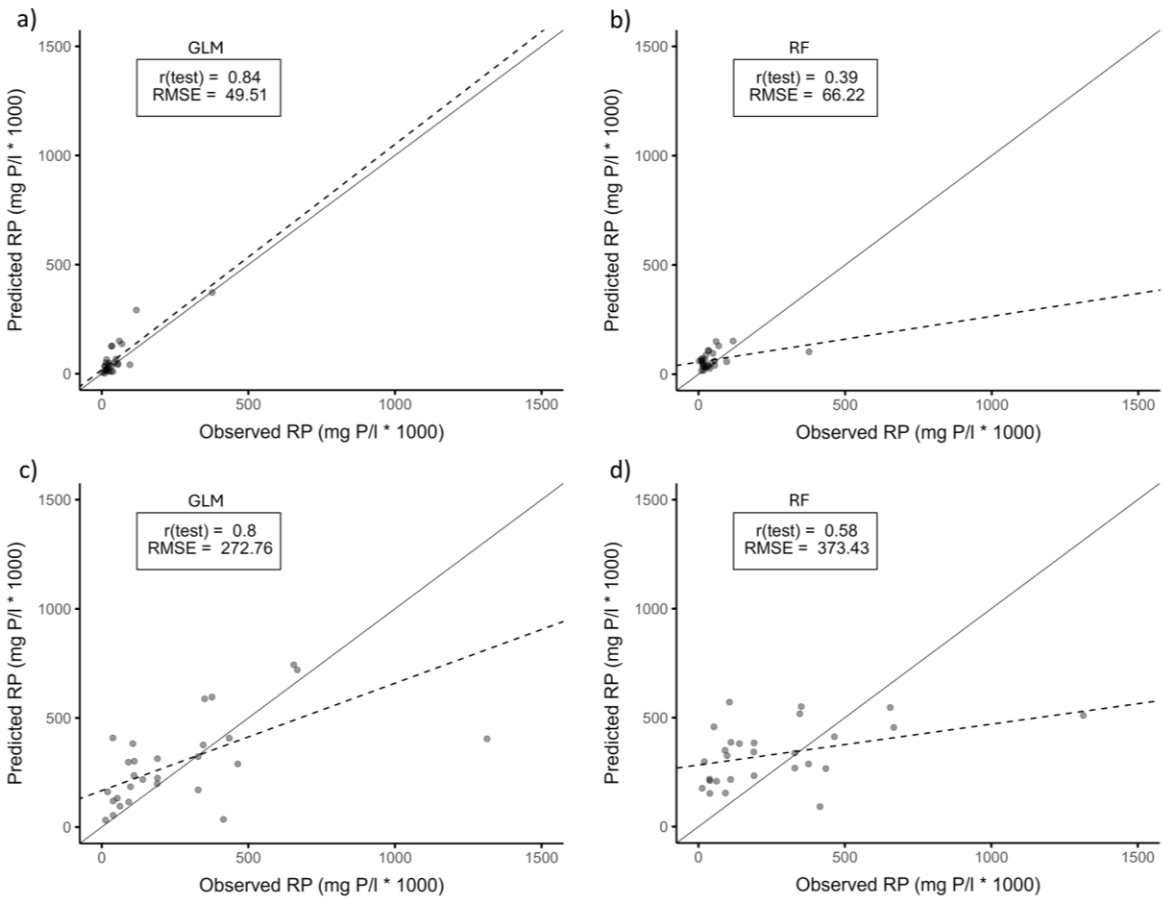


Figure 8: Correlation plots for the test data (n=29) for the negative binomial generalised linear model (a) and random forest model (b) of RP in catchments with low population density, and the negative binomial generalised linear model (c) and random forest model (d) of RP in catchments with high population density. Horizontal axes show the true values from the test data set, multiplied by 1000 in order to obtain an integer, whilst the vertical axes show the values predicted by the model. The dashed line shows the linear regression of the data points and the solid line represents the 1:1 relationship. The box in the upper left corner gives the Pearson coefficient and the value for RMSE for each model. The labels GLM and RF refer to generalised linear model and random forest respectively.

3.3.2 Effect size and variable importance

As expected under (H1), the generalised linear model for TON shows that agricultural sources, namely arable and horticultural land cover, and cattle density, are significant positive predictors of TON in catchments with low population density, whilst the population equivalent of WWTWs is not a significant predictor of TON in these catchments (Table 5a). This is confirmed by the random forest models of TON in catchments with low population density, as it ranks arable and horticultural land use as one of the top two predictors of TON (with the other predictor being the HOST base flow index) (Figure 9).

The results from the models of RP in catchments with low population density lend some support to (H1), as arable and horticultural land use, and cattle density, are both significant positive predictors of RP in the negative binomial generalised linear model of catchments with low population density (Table 5b), and have a bigger effect size than the population equivalent of WWTWs in the same model (Table 5b, and Figure A5.1 in Appendix 5). However, the population equivalent of WWTWs is still a significant predictor in catchments with low population density, and sheep density has a negative effect. The results from the random forest model of RP in catchments with low population density show that arable and horticultural land use and cattle density rank higher than the population equivalent from WWTWs (Figure 9), however the low R^2 for this random forest model, as well as the evidence of overfitting (Figure 8b), means that these results should be interpreted with caution.

Table 5: *Formula, estimated regression parameters, standard errors, z-values and p-values for the minimum adequate negative binomial generalised linear models of catchments with low population density for the TON data (a) and RP data (b), and of catchments with high population density for the TON data (c) and RP data (d). Model R^2 are 0.41, 0.47, 0.35, and 0.61 respectively.*

(a)

$TON_{low\ pop}$ ~ ArableHortProp + MeanSlope + LogCatchmentArea + LogCattleDensity + PopDensity + LogMaxPrecipitation + HOSTBaseFlowIndex + LogDesignatedAreaProp

	Estimate	Std. error	z value	p-value
Intercept	7.55	0.04	180.35	< 0.001
ArableHortProp	0.35	0.08	4.52	< 0.001
MeanSlope	0.26	0.08	3.22	< 0.05
LogCatchmentArea	0.09	0.05	1.79	0.073
LogCattleDensity	0.30	0.05	6.17	< 0.001
PopDensity	0.22	0.05	4.07	< 0.001
LogMaxPrecipitation	-0.57	0.08	-6.94	< 0.001
HOSTBaseFlowIndex	0.36	0.05	7.08	< 0.001
LogDesignatedAreaProp	-0.15	0.06	-2.29	< 0.05

(b)

$RP_{low\ pop}$ ~ LogForestProp + ArableHortProp + ChannelDensity + LogCatchmentArea + LogCattleDensity + LogSheepDensity + PopDensity + LogPopEquiWWTW

	Estimate	Std. error	z value	p-value
Intercept	3.68	0.07	51.11	< 0.001
LogForestProp	0.17	0.10	1.74	0.081
ArableHortProp	0.31	0.10	3.09	< 0.01
ChannelDensity	0.19	0.10	2.01	< 0.05
LogCatchmentArea	-0.49	0.11	-4.68	< 0.001
LogCattleDensity	0.45	0.11	4.30	< 0.001
LogSheepDensity	-0.33	0.14	-2.40	< 0.05
PopDensity	0.33	0.11	2.94	< 0.01
LogPopEquiWWTW	0.30	0.09	3.34	< 0.001

(c)

$TON_{high\ pop}$ ~ MeanSlope + ChannelDensity + LogCatchmentArea + LogCattleDensity + HOSTBaseFlowIndex + LogPopEquiWWTW

	Estimate	Std. error	z value	p-value
Intercept	8.71	0.05	188.26	< 0.001
MeanSlope	-0.25	0.05	-4.77	< 0.001
ChannelDensity	-0.13	0.05	-2.50	< 0.05
LogCatchmentArea	-0.24	0.07	-3.55	< 0.001
LogCattleDensity	0.08	0.05	1.47	0.141
HOSTBaseFlowIndex	0.13	0.05	2.61	< 0.01
LogPopEquiWWTW	0.52	0.07	7.62	< 0.001

(d)

$RP_{high\ pop}$ ~ ArableHortProp + LogCatchmentArea + LogMaxPrecipitation + HOSTBaseFlowIndex + LogPopEquiWWTW

	Estimate	Std. error	z value	p-value
Intercept	5.65	0.09	65.63	< 0.001
ArableHortProp	-0.25	0.11	-2.32	< 0.05
LogCatchmentArea	-0.57	0.12	-4.69	< 0.001
LogMaxPrecipitation	-0.31	0.11	-2.92	< 0.01
HOSTBaseFlowIndex	-0.35	0.09	-3.78	< 0.001
LogPopEquiWWTW	1.06	0.12	9.28	< 0.001

The independent variables are the proportion of arable and horticultural land cover; channel density; the HOST base flow index for the catchment; log of the catchment area; log of the cattle density; log of the proportion of the catchment designated for recreation or nature conservation; log of the proportion of forest land cover; log of maximum average yearly precipitation; log of the population equivalent for all the WWTWs in the catchment; log of the sheep density; mean slope; population density.

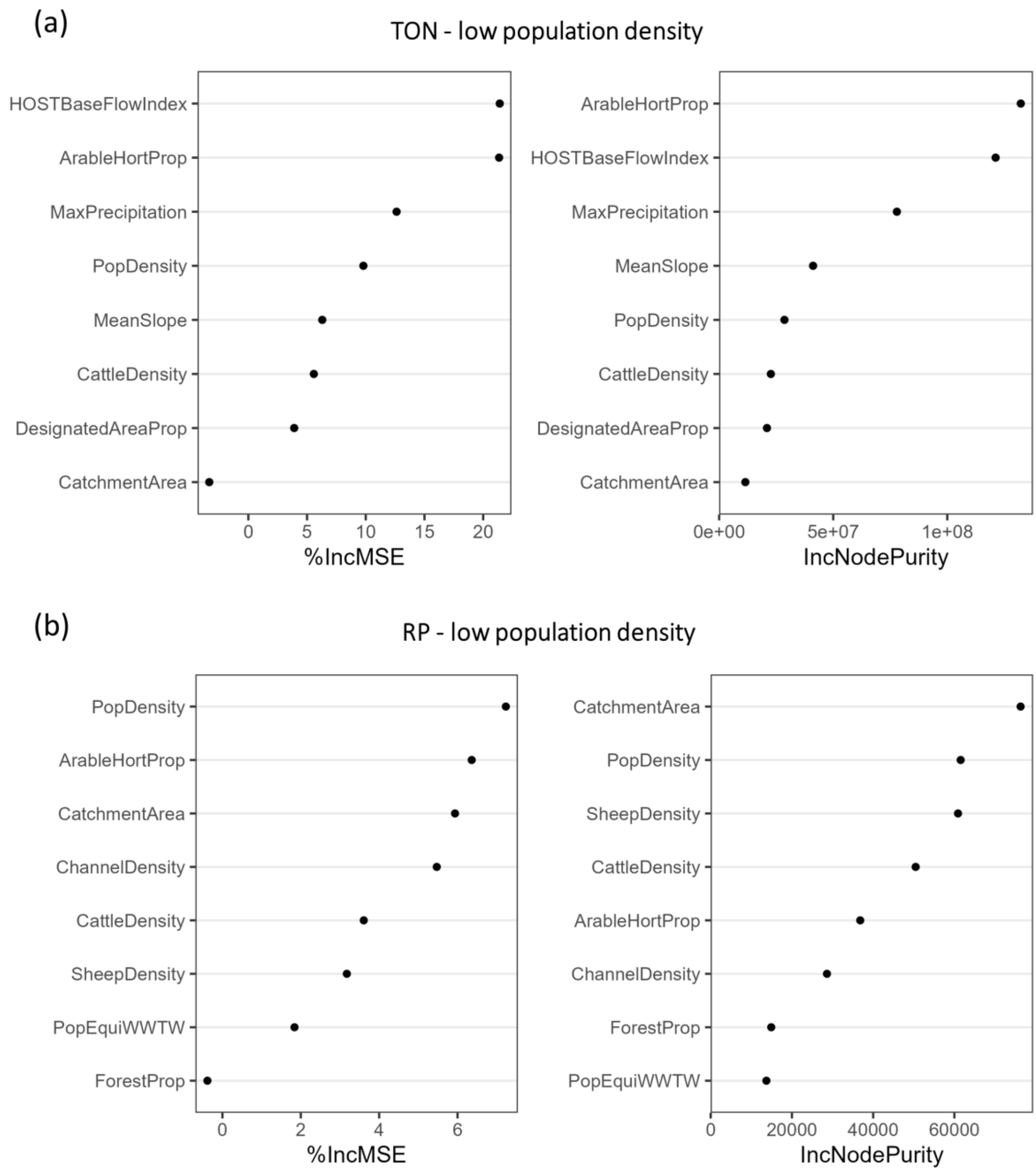


Figure 9: Variable importance for the random forest for TON in catchments with low population density (a), RP in catchments with low population density (b). In each case the left hand panel shows the permutation importance and the right hand panel the Gini importance.

As expected under (H2), the generalised linear model for TON in catchments with high population density shows that arable and horticultural land cover and sheep density are not significant predictors of TON, whilst the population equivalent of WWTWs is a significant positive predictor of TON (Table 5c), with a large effect size (Table 5c, Figure A5.2a in Appendix 5). The random forest model for TON in catchments with high population density also supports (H2), with the population equivalent of WWTWs ranking as the top predictor of TON (Figure 10).

The results for RP in catchments with high population density also support (H2), as arable and horticultural land cover is a negative predictor of RP (Table 5d), while the population equivalent of WWTWs has a comparatively large positive effect on RP concentrations in these catchments (Table 5d, Figure A5.2c in Appendix 5). The random forest for RP in catchments with high population density confirmed this, with the population equivalent of WWTWs ranking as one of the top predictors (with the other being catchment area) (Figure 10).

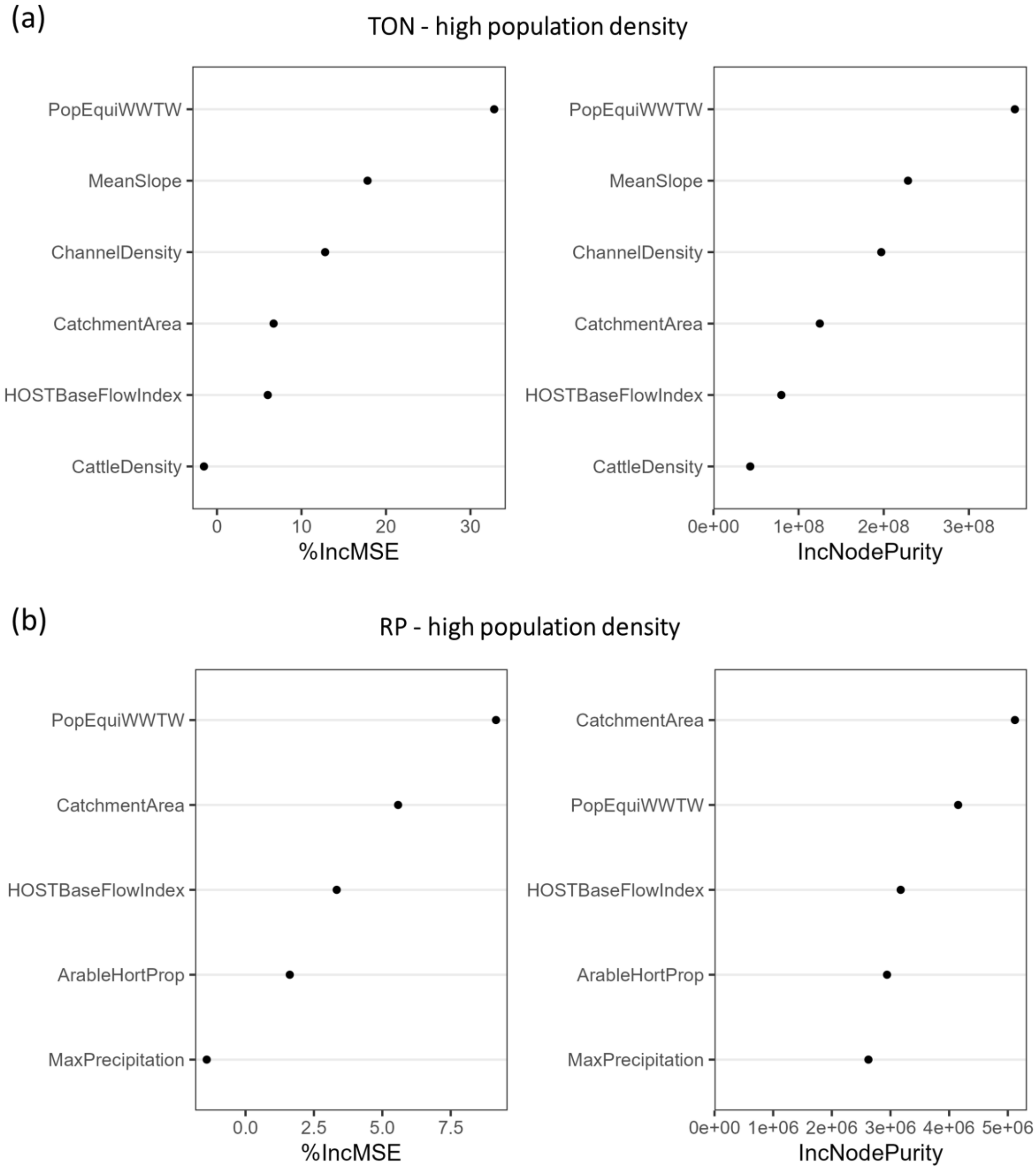


Figure 10: Variable importance for the random forest for TON in catchments with high population density (a) and RP in catchments with high population density (b). In each case the left hand panel shows the permutation importance and the right hand panel the Gini importance.

3.4 Discussion

In this study we use statistical techniques to model N and P concentrations nationally in England, comparing the results from catchments with low population density to catchments with high population density, and demonstrate how these techniques can be used to understand the relative importance of different sources of N and P in a way that is relevant to management and policy. Our models for N and P show satisfactory predictive ability for the most part, showing there is potential to use this approach more widely, although they consistently underestimate very high mean TON and RP concentrations (approx. > 12 mg/l N and approx. > 0.8 mg/l P), and the model validation results were poor for some of the models of RP. With regards to our first hypothesis, our results show that agricultural sources of N dominate in catchments with low population density in England, as was expected, and this is true to some extent for P. In terms of management and policy, this is important because it suggests that an emphasis on agriculture in low population watersheds is needed to mitigate nutrient impairment, although of course reductions of inputs from small WWTWs (and septic tanks) will also be beneficial. Our results also confirm our second hypotheses, as they show that in catchments with high population density, WWTWs are the most important sources of N and P, and thus management efforts in these catchments should prioritise reducing inputs from these point sources. These results lend support to previous suggestions (e.g. Withers et al. 2014) that the contribution of agricultural sources may be overestimated in catchments that are densely populated, relative to point sources from waste water treatment works, as the later dominates as a predictor of N and P concentrations in these catchments. This study has shown that agricultural sources of N and P are *comparatively* more important in catchments with lower population density. The mean concentrations of TON and RP are lower in the catchments with lower population density, so WWTWs could have been just as important in determining these concentrations as the WWTWs in the higher population density catchments (because the concentrations are lower and therefore, even though the

WWTWs are smaller, their relative effect could have been the same or greater). It seems likely that as population increases so do other sources of N and P that were not directly included in the model – such as runoff from roads and urban areas, industrial effluent, illegal discharges and septic tanks– which could cumulatively become an important source of N and/or P, and thus makes agricultural sources comparatively less important as a predictor in these catchments. There is a slight correlation between population density and the population equivalent from WWTWs (Pearson's correlation < 0.37 for all datasets), as would be expected, and this will go some way towards explaining why the population equivalent of WWTWs is an important predictor of N and P concentration in watersheds with high population density. Previous studies have suggested that P enrichment is more likely to be the cause of nutrient impairment in lowland, high alkalinity rivers (Jarvie et al. 2018), and with this in mind, a continued emphasis on RP reduction at WWTWs in urban areas is likely to be the right approach to improving water quality in catchments with high population density.

The findings of this study have implications for the debate around the relative importance of agricultural diffuse sources and point sources from WWTWs to nutrient concentrations in rivers, and help understand which management strategies are most likely to be effective in different contexts. However, the relationships we found are a generalisation based on a large-scale assessment, so for local decision making many other aspects will be relevant, including the local environmental conditions, and social and economic aspects. Moreover, this approach is limited to making recommendations around large-scale land cover and land management (e.g. livestock density) changes. Quantification of the overall effectiveness of smaller scale measures, for example buffer strips, contour ploughing, at the national, or catchment, scale remain elusive.

There are also several limitations to the data used in the study. Firstly, the geographical distribution of the catchments with low population and high population density used in this study are not identical (as seen previously in Figure 6), which means we cannot completely rule out that the signal being picked up is due to some other variable that varies regionally and is not accounted for by the model, such as the main crop type in arable areas, or the distribution of industry. We also did not have access to the type of treatment applied to waste water at the different WWTWs, nor the exact population equivalent for the smaller WWTWs (those with population equivalent < 2000). Access to this information at a national scale would help provide a more nuanced picture of the contribution of WWTWs to N and P in freshwater ecosystems.

There are other considerations when interpreting the results. Firstly, the findings of this study cannot be applied to catchments with characteristics outside of the range available for inclusion in this study, such as upland catchments. Secondly, a limitation of this study is that the models are not giving information about particulate transport of N and P, and not separating between organic and inorganic forms. This means that the results are only able to present a partial picture of N and P retention, and this is likely to be particularly important for P, as the particulate transport pathways are known to be important (Reaney et al. 2011). However, the use of TON and RP means that the focus is on the predominant forms that affect plant growth, as they are readily available for uptake (Prasad and Chakraborty 2019, GRDC 2013). There is also not currently enough data available on Total N and Total P concentrations, or organic N and P, at monitoring stations in England to use the approach presented in this study on these deteminands, but this could be an interesting avenue of future study. Thirdly, there is the question of spatial configuration of the catchment characteristics. The models give insight into the importance of various catchment characteristics which have been summarised at a catchment level, in generally large catchments, but the situation may be very different at a local scale. Certain catchments characteristics, such as the proportion of area covered by forest, were not generally shown to be important to predict TON and RP in this study, but they may or may not play a role more locally in patches or as buffer zones along a river. Finally, defining the catchment for each of the monitoring stations in an automated way was challenging, and although we carried out a large quality-control effort through visual inspection and comparison with other available datasets, it is possible that some mistakes remain in catchment definition, which would then affect all independent variables for that monitoring station.

Much progress has been made to better understand the sources and dynamics of natural and anthropogenic inputs of N and P into rivers (e.g. Jarvie et al. 2018, Johnes et al. 2022). However, modelling N and P in rivers at large scales remains challenging, and different approaches have emerged to tackle the problem, including empirical models, such as the export coefficient models (e.g. Redhead et al. 2018, Johnes et al. 1996), as well as processed based models (e.g. the LTLS Freshwater Model described in Bell et al. 2021). The increasing availability of large and often publicly available datasets with water quality measurements and other environmental data has led to an increase in the use of statistical techniques to study water quality (Schreiber et al. 2022, Spake et al. 2019, Moorhouse et al. 2018, Tate et al. 2003), as we have done in this study. These techniques are in the spirit of a wider body of work that aims to develop data science and artificial intelligence techniques for the natural environment (Blair 2021, Scowen et al. 2021, Breiman et al. 2001, Lucas 2020), in the hope that environmental science and ecology can reap the benefit of the increasing quantity and diversity of data available to researchers. However, the approach we used has limitations, including inconsistent results between different measures of variable importance in some cases, and results that may not make sense from a process perspective, such as the negative effect of arable and horticultural land use on RP concentrations in the generalised linear model for catchments with high population density. Overall, however, the results from the generalised linear models and random forest models were fairly consistent with each other for each dataset, which is reassuring. The model validation results do, however, highlight that outliers have a strong effect on the models and that particularly random forest tended to overfit in these situations.

An interesting avenue going forward would be to use the models in this study to make predictions, that would allow a more nuanced discussion of the decrease in nutrient concentration that could be expected to result if land cover change were to take place in different contexts. In addition to this, adapting the approach to take seasonality into account in some way is probably important to be able to discuss to what extent the findings of this study are ecologically meaningful, as variable importance may well vary seasonally and the effect of changes to catchment characteristics may also be sensitive to seasonality. For example, it would help to consider ecologically sensitive periods in spring and summer when rooted aquatic plants and algae grow (Jarvie et al. 2006). Finally, conducting this type of study in other countries or regions would help understand to what extent the findings of this study are specific to England or represent more general patterns due to the nature of N and P transport and retention.

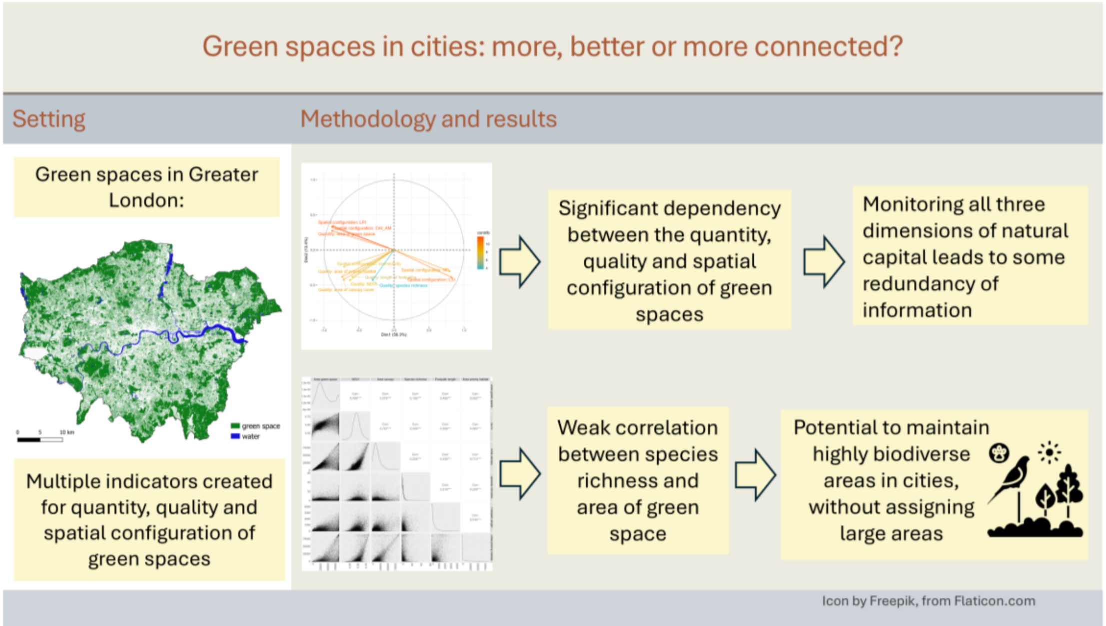
3.5 Conclusion

The sources and dynamics of natural and anthropogenic inputs of N and P into rivers remains a complex problem, and despite substantial domain knowledge about these processes it remains challenging to model. However, understanding the relative importance of diffuse sources from agriculture, and how this varies in different contexts, is important because it allows us to spatially target management strategies to the places where they are likely to have the strongest positive effect, and through this balance the provision of food production and good water quality. The results of this study need to be interpreted with some caution, but they do provide some insight and recommendations. Firstly, our results suggest that management strategies aimed at reducing N and P from agricultural sources might be better suited to catchments with low population density, although local factors would of course also be important in any decision-making process. Secondly, they suggest that to reduce the concentration of TON and RP in catchments in England with high population density, a continued focus on WWTWs as point sources should be a priority. Going forward, more could be done to make detailed data on WWTWs available, including their population equivalent and the type of treatments applied, which would make it easier to include this independent variable in all types of models.

Climate change is likely to increase pressure on river systems and the benefits they support, which will increase the need to target management strategies to preserve the multiple benefits we receive from nature. The debate about the relative contribution of diffuse agricultural sources and point sources from WWTWs to N and P concentrations in rivers will only become more relevant, as these two sources are affected by climate change in different ways (Wade et al. 2022). One way of furthering our understanding of these processes is through harnessing the opportunities brought about by the increasing availability of diverse environmental datasets (Lavallin and Downs 2021, Blair and Henrys 2023), and the development of methods and approaches to use these data to gain insight (Yu et al. 2021). This study explored a particular approach to this, using well established methods and a broad range of environmental data, highlighting some of the opportunities and challenges in the approach. Exploring new opportunities in this area is particularly important because biogeochemical flows (N and P) are one of the six planetary boundaries that have been found currently to be transgressed, suggesting that they are contributing to the Earth being outside of the safe operating space for humanity (Richardson et al. 2023).

Chapter 4: Green spaces in cities: more, better or more connected?

Visual Abstract



Abstract

Urban green spaces provide a broad range of benefits to society, making them an important element of natural capital accounting. Environmental accounting efforts aim to capture not only the quantity of green space in cities, but also their quality and spatial configuration, as these all influence the benefits we receive from them. So far, few studies have explored the dependency between these three characteristics or "dimensions" of natural capital, even though it has implications for both monitoring efforts and management. To address this knowledge gap, we here assess the correlation structure between different indicators for the quantity, quality and spatial configuration of green spaces, using Greater London as a case study. Our results show that all indicators show significant dependency with each other, suggesting it may not always be necessary to monitor all three dimensions of natural capital. However, the strength of dependency was shown to vary between different indicators, with spatial configuration and quantity generally displaying strong positive correlation. Species richness, an indicator of quality, showed the weakest correlation with the quantity of green area; this suggests that to create highly biodiverse areas in urban settings, simply extending the area of habitat may not be enough and more targeted management is likely to be necessary.

4.1 Introduction

Currently more than 50% of the world's population lives in urban areas (United Nations 2020), which means that a large proportion of people's experience of the natural world occurs in towns and cities, and many of the benefits that people receive from nature occur in an urban context (O'Keeffe et al. 2022, Hamel et al. 2021). There is growing awareness of the importance of nature in cities (Mata et al. 2020), and a call for more and better management of nature in cities has led to an uptake of natural capital accounting for urban areas (e.g. Phinney 2022, Mayor of London 2017, Yang et al. 2021). Natural capital is another term for the world's stock of natural assets, which include renewable and non-renewable natural resources (e.g. plants, animals, air, minerals, freshwaters) that combine to yield a flow of benefits to people (Mace et al. 2015). Natural capital accounting is a term covering efforts to use an accounting framework to measure and report on natural capital assets and the flow of benefits we receive from it in a systematic way (SEEA 2021). The growth of cities has led to major changes in the amount, quality and distribution of natural habitats within urban areas and the services they provide to society, putting some benefits at risk (Gómez-Baggethun and Barton 2013). The natural capital framework has emerged as an approach to monitor these changes, often with the aim of informing management efforts to "future-proof" urban areas against the challenges of climate change and rapid urbanisation (United Nations 2023).

Urban green space in cities is a particularly important element of natural capital accounting for urban areas, due to the range of benefits we receive from them (Phinney 2022). Many of the benefits that people receive from green spaces are environmental. For example, green spaces can help reduce the heat gain of cities and lessen the negative impacts of heatwaves on human health (Heaviside et al. 2017). They can also offer opportunities for carbon sequestration (Wang et al. 2023), especially if green spaces are managed to increase tree cover (Ariluoma 2021). Green spaces have moreover been shown to reduce air pollution concentrations (Jones et al. 2019) and reduce the risk of floods, by increasing the interception and storage capacity of the urban landscape, and thus reducing storm water runoff (Zimmermann et al. 2016, Campbell et al. 2020). Other values of green space reflect their use for recreation, with individuals, public services and business all benefiting from public parks across cities, as they create opportunities to exercise, socialise and relax, improving peoples physical and mental health (van den Bosch and Sang 2017). Whilst many of these benefits flow from large public parks, private gardens also have an important role to play, as garden use is associated with wellbeing, physical activity, and visiting nature (de Bell 2020). Private gardens can also provide refuge for wildlife (García-Antúnez et al. 2023) and increase habitat connectivity between larger green spaces (Rudd et al. 2002, Hanson et al. 2021).

Accounting for green space within a natural capital framework involves reporting on the different characteristics of green spaces that determine the benefits we receive from them. These characteristics or "dimensions" of natural capital are the quantity, quality and spatial configuration of assets, as all of these have implications for the benefits we receive (Mace et al. 2015). Quantity refers to the "amount" of the asset, in this case the amount of green space. The "quality" of green space refers to its condition, and it will be particularly important if the presence or absence of a certain component or process affects the benefits we receive from green spaces (Mace et al. 2015), for example a degraded park may be less enjoyable to visit and will provide limited carbon sequestration. Finally, the spatial configuration of an asset refers to the spatial patterning and fragmentation of the asset, which can influence the benefits that flow from green spaces. For example, the connections (including proximity) between green spaces facilitate metapopulation dynamics between them (Hanski 2015) and decreases the heat island effect (Li and Zhou 2019, Chen et al. 2014, Kong et al. 2014), whilst evenly distributed green space throughout a city means that a higher proportion of the population will have easy access to them and be able to enjoy the recreational benefits they provide (Handley et al. 2003).

Monitoring the quantity, quality and spatial configuration of urban green space is important within urban natural capital accounting, and existing work uses, for example, the extent of green space and broad habitats as measures of quantity (Office for National Statistics 2023b), compositional species indicators and urban trees as measures of quality (Office for National Statistics 2023b), and patch size, shape and edge as measures of spatial configuration (Natural England 2018). Considering all three dimensions of natural capital leads to a substantial effort in terms of gathering data, developing indicators for the different dimensions of natural capital, and reporting on them. As an approach it treats the quantity, quality and spatial configuration of assets as implicitly "orthogonal", that is, conceptually as if mapping out asset condition as a three dimensional space, in which all dimensions are needed to describe asset status. It is not clear, however, whether the indicators of these three dimensions are systematically mutually independent in this way. For example, there is evidence of dependency between measures of quantity and some common measures of spatial configuration, namely landscape metrics (Neel et al. 2004, Wang et al. 2014). This means that there may be some redundancy in the information provided by the indicators for the three dimensions of natural capital. Any dependency between the three dimensions of natural capital may also have implications for management, as if they are closely related it should be possible to develop approaches that boost more than one dimension.

However, to our knowledge, no study has tested the degree of dependency between indicators of all dimensions of natural capital assets (namely their quantity, quality, and spatial configuration), despite the relevance to monitoring efforts and management. To fill this gap in knowledge, we assess the degree of dependency that exists between the quantity, quality and spatial configuration of green spaces in the largest and most populated capital in western Europe, namely London in the United Kingdom (UK). We aim to test the following hypotheses: Firstly, based on previous works, we expect dependency between the three dimensions of natural capital assets (Jaganmohan et al. 2016, Smith et al. 2009) (H1). Secondly, we expect the dependency between quantity and spatial configuration to be the strongest, as this was previously reported for landscape metrics and habitat abundance (Neel et al. 2004, Wang et al. 2014) (H2). Thirdly, the strength of the correlation will vary depending on the indicator chosen for each of the three dimensions (Neel et al. 2004, Wang et al. 2014) (H3). Finally, we expect the dependency between the three dimensions of natural capital assets to persist at different scales of analysis (H4).

4.2 Materials and methods

4.2.1 Study area

Greater London is the administrative area of London, the capital of the UK, covering an area of 1,569 km² and situated in the south-east of England (Figure 11). It has a temperate oceanic climate, with mild winters and temperate summers. Precipitation is fairly evenly distributed throughout the year, with a total annual precipitation of 585 mm (Britannica 2023). London's mid-2021 population was 8.797 million (GLA Data Store 2023) and it is considered one of the greenest cities in the world for its size, although this greenspace is unequally distributed (Greenspace Information for Greater London CIC 2023a).

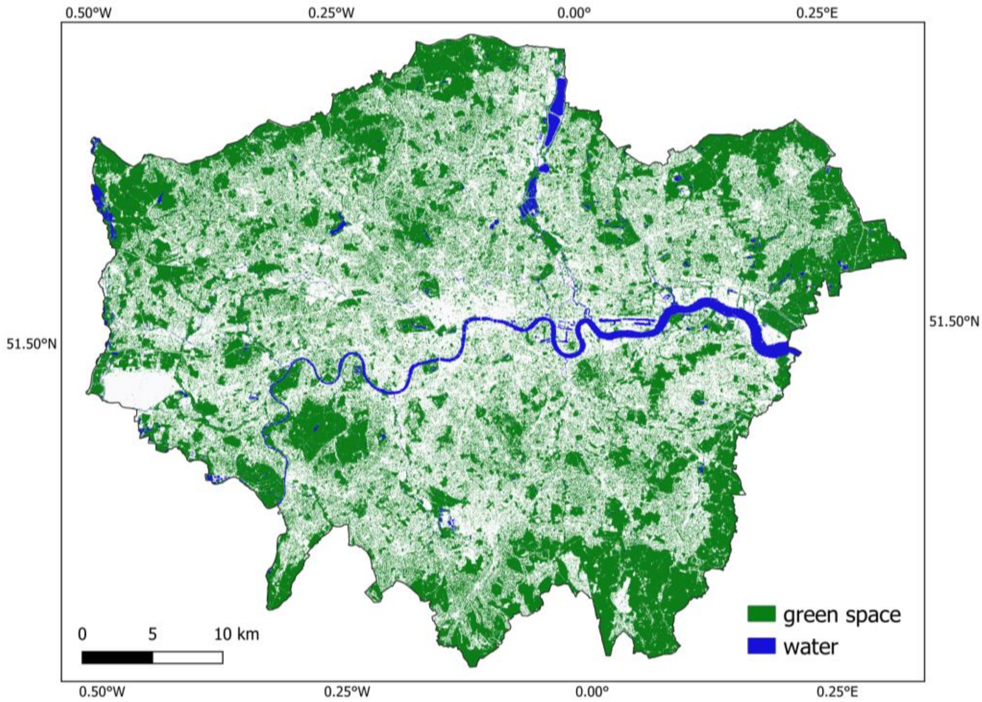


Figure 11: Map of Greater London showing green spaces and large water bodies. Data from Greater London Authority's Green and Blue Cover dataset (GLA City Intelligence 2019).

Greater London comprises more than 47% greenspace (London Councils, 2018); 33% of London is natural habitats within open space according to surveyed habitat information (Greenspace Information for Greater London CIC 2023b) and an additional 14% is estimated to be vegetated private, domestic garden land (Smith et al. 2011). The median garden size for a house in London is 140 m^2, and one in five households in London do not have a private garden (Office for National Statistics 2020). Priority habitats in London's open green space includes woodland, acid grassland, chalk grassland and heathland (Greenspace Information for Greater London CIC 2009). Most public green spaces are managed by the London Boroughs, other public agencies (such as The Royal Parks and Lea Valley Regional Park Authority) and environmental organisations (such as London Wildlife Trust).

4.2.2 Data acquisition

We created indicators for the quantity, quality and spatial configuration of green spaces (Table 6) across Greater London at 300 m resolution. We chose eleven indicators in total, including indicators that are commonly used within natural capital reporting, and other that are logically consistent with the natural capital concept and have a clear link to benefit provision. We included five indicators for quality, five indicators for spatial configuration, and only one indicator for quantity (total green space area), as it is the only indicator of the quantity of green space that we found being used in the literature. To calculate the indicators at 300 m resolution, we created a grid across Greater London with each square 300 by 300 m, and assigned a value for each indicator to each cell in the grid. We chose this spatial resolution as our primary scale of analysis as it is the finest resolution that can be used meaningfully with the coarser resolution datasets, especially the species data, and a fine scale resolution allows gardens and smaller green areas to be considered (Myint et al 2011, Crowson et al. 2024). All data preparation steps were carried out in R (R Core Team 2022) unless otherwise stated.

Table 6: *Indicators used in this study for the quantity, quality and spatial configuration of green spaces in Greater London.*

Dimension	Indicator	Definition
Quantity	Total area of green space	Amount of green space in a given 300 x 300 m grid.
Quality	Species richness	Number of species in a given 300 x 300 m grid.
Quality	Normalized difference vegetation index (NDVI)	Mean NDVI of green spaces in a given 300 x 300 m grid
Quality	Total area of tree canopy cover	Area covered by trees in a given 300 x 300 m grid.
Quality	Total length of footpaths	Length of footpath in a given 300 x 300 m grid
Quality	Total area classified as a habitat of principal importance	Area of green space classed as a habitat of principal importance in a given 300 x 300 m grid.
Spatial configuration	Area weighted mean core area index (CAI_AM)	CAI_AM for green spaces in a given 300 x 300 m grid, calculated as a class metric. The metric CAI_AM quantifies core area of green space as a percentage of total green space area.
Spatial configuration	Landscape shape index (LSI)	LSI for green spaces in a given 300 x 300 m grid, calculated as a class metric. The metric LSI is the ratio between the actual edge length of green space and the hypothetical minimum edge, if green space were maximally aggregated.
Spatial configuration	Largest patch index (LPI)	LPI for green spaces in a given 300 x 300 m grid, calculated as a class metric. The metric LPI is the percentage of the grid covered by the corresponding largest patch of green space.
Spatial configuration	Number of patches (NP)	Number of patches of green space in a given 300 x 300 m grid, calculated as a class metric.
Spatial configuration	Landscape connectivity	Mean omni-directional landscape connectivity for green spaces in a given 300 x 300 m grid.

For the indicator of quantity of green space in Greater London we calculated the total area of green space in each 300 by 300 m cell. The total area or extent of a habitat is a common indicator for the quantity of an asset (Natural England 2018), including for natural capital assessments of green spaces in an urban setting (Office for National Statistics 2023b, Natural England 2021, Phinney 2022). To calculate the area of green space we used the Greater London Authority's Green Cover dataset (GLA City Intelligence 2019) which maps vegetated land cover in London (as opposed to man-made surfaces), based on high resolution (25 cm) near-infrared aerial imagery. This dataset is provided as a vector, so we rasterized the dataset at the original 25 cm resolution using the terra package (Hijmans 2022), and to reduce the error in the data and make it consistent with the other indicators we masked the data to include only areas identified as open space or gardens, identified by the London-wide database of open spaces (Greenspace Information for Greater London CIC 2023c) and the OS MasterMap Greenspace Layer (Ordnance Survey Limited 2022). From this we calculated the total vegetated area within each 300 x 300 m cell across Greater London.

We chose five indicators for the quality of green space, (1) species richness, (2) Normalized Differential Vegetation Index (NDVI), (3) extent of tree canopy cover, (4) length of footpaths and (5) extent of habitats of principal importance. Species richness was chosen to capture highly biodiverse areas, which can be considered high quality areas (Natural England 2018). NDVI is an indicator of vegetation productivity derived from remote sensing data (Pettorelli, 2013), and has widely been used to reflect the vegetation quality in urban areas (Wang et al. 2019, Sarkar et al. 2015), especially in applications that study the urban thermal environment (Jaganmohan et al. 2016; Yang et al. 2017). NDVI can also be an indicator of the capacity of urban green space to provide certain ecosystem services, such as carbon sequestration, and water cycling (Zurlini et al. 2014). We chose the extent of tree canopy cover as an indicator of the quality of green space because the number or density of trees has been used to represent the quality of greens space in the past (Sarkar et al. 2015, Zhou et al. 2017), and trees are strongly related to many of the benefits we receive from green space, as they intercept rain, clean the air, and provide shade and habitat. The total length of footpaths in greens spaces has previously been used as an indicator of the quality of green spaces in urban environment (eftec 2017), because it relates to the accessibility of green spaces for people, which in turn is important for the provisions of opportunities for sport and leisure. The total area of green spaces classified as Natural Environment and Rural Communities Act (2006) Section 41 habitats of principal importance has been used previously in natural capital assessments (Holt 2017) to capture habitats that are at risk or are of importance for key species, and are thus providing important biodiversity conservation benefits (Maddock 2011).

Species richness was calculated using the species records for London's Priority Species List (Greenspace Information for Greater London CIC 2023d), which includes species that are national priorities for conservation and those that are believed to be declining in London or beyond. The species occurrence records are from a range of dates, collected by different organisations in different ways. We included all terrestrial species that are relevant to green spaces, including birds, terrestrial mammals, reptiles, invertebrates, plants and fungi. To create the indicator of species richness, we added up the number of different species in each 300 by 300 m grid.

To calculate NDVI we used PlanetScope images of Greater London from the 7th of August 2022 (3 m resolution, Planet Labs PBC 2022). The date was chosen as there was full coverage of our study area with 0% cloud cover, and deciduous trees had their leaves on. The NDVI is a vegetation index derived from the red (RED) to near-infrared (NIR) reflectance ratio, calculated as NDVI = (NIR – RED)/(NIR + RED). The resulting values range from -1 to +1, with green leaves and vegetation resulting in positive NDVI values, bare soil and concrete resulting in NDVI values close to zero, and water resulting in negative NDVI values (Pettorelli 2013). For each 300 by 300 m grid, we took an average of the NDVI value for pixels within green spaces (defined using the Greater London Authority's Green Cover dataset, used previously to create the indicator of quantity).

To calculate the third indicator for quality, the area of canopy cover within green areas, we used an existing canopy map of Greater London, created at 25 cm resolution using colour infrared imagery (Breadboard Labs 2018). To calculate the fourth indicator of quality, total length of footpaths within green spaces, we used OpenStreetMap (2024) data on "paths unsuitable for cars", which includes footpaths, paths for horse riding and paths for cycling. Finally, to calculate the total green area classified as a habitat of principal importance we used Natural England's Priority Habitat Inventory V2.3 (Natural England 2024), and included all terrestrial habitats found within green spaces.

We used five indicators for the spatial configuration of green space, (1) Area Weighted Mean Core Area Index (CAI_AM) (McGarigal et al. 2012), (2) Landscape Shape Index (LSI) (McGarigal et al. 2023), (3) Largest Patch Index (LPI), (4) Number of patches (NP) and (5) omni-directional landscape connectivity (Landau et al. 2021). Landscape metrics are one of the most common metrics for spatial configuration used within the Natural Capital framework (Natural England 2018) and we chose to use CAI_AM, LSI, LPI and NP because they capture different aspects of patches of green space (core area, aggregation, area/edge, and shape respectively), and existing work suggests that CAI_AM should not correlate as strongly with habitat extent as some other landscape metrics (Neel et al. 2004, Wang et al. 2014). In addition to the landscape metrics, we included an indicator for landscape connectivity, as this is an important aspect of spatial configuration and is central to the UK Government's target of making areas of semi-natural habitat "more, bigger, better and joined up" (Mancini et al. 2022). We chose to use the connectivity measure from Omniscape.jl because it does not require the user to divide the landscape into core areas to be connected and because the algorithm it implements calculates connectivity in all directions, which sets it apart from other similar methods that require the user to set a direction across which to measure connectivity (for example north to south) (McRae et al. 2016).

We calculated CAI_AM, LSI, LPI and NP as class metrics in Fragstats 4.2 (McGarigal et al. 2012) for each 300 m grid cell, with green space as the class of interest, based on the Greater London Authority's Green Cover data (GLA City Intelligence 2019) aggregated to 3 m resolution (Figure 12). The metric CAI_AM quantifies core area of green space as a percentage of total green space area, and edge depth was fixed at 5 m based on the resolution of the underlying data and previous studies (Karimi et al. 2021, Grafius et al. 2018). The metric LSI is calculated as the ratio between the actual edge length of green space and the hypothetical minimum edge, if green space were maximally aggregated. The metric LPI is the percentage of each 300 m x 300 m grid covered by the corresponding largest patch of green space, and is a simple measure of dominance. The metric NP is the number of patches of green space in each 300 m by 300 m grid (McGarigal et al. 2012).

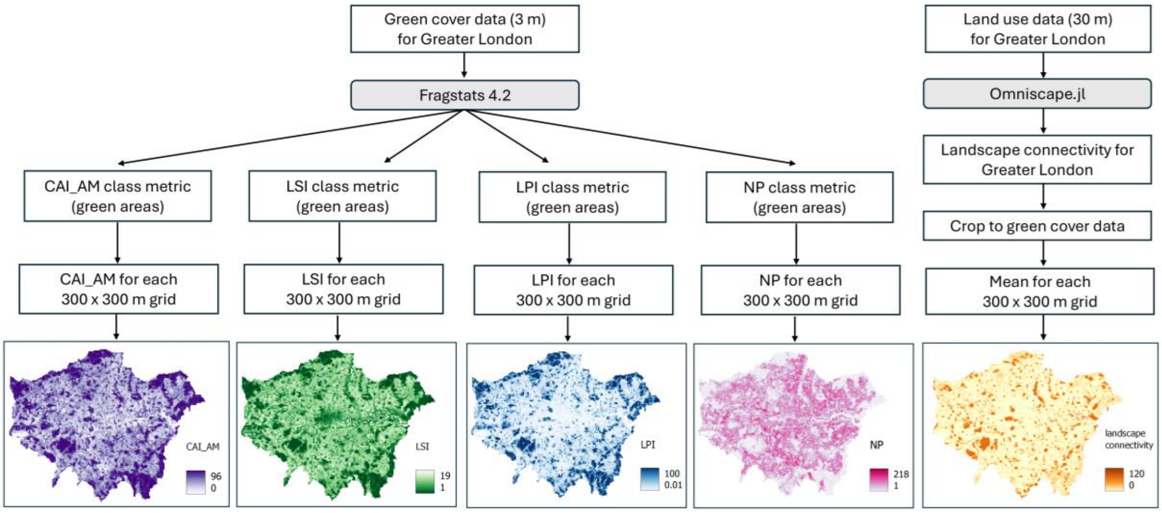


Figure 12: Generalised work flow for the metrics of spatial configuration used in this study, namely the landscape metrics Area Weighted Mean Core Area Index (CAI_AM), Landscape Shape Index (LSI), Largest Patch Index (LPI), Number of Patches (NP) and omni-directional landscape connectivity. The map of landscape connectivity was calculated using data on land use provided by Greenspace Information for Greater London CIC 2023c.

To create an indicator for landscape connectivity we used the software package Omniscape.jl (Landau et al. 2021, McRae et al. 2016), which uses circuit theoretic methods to model ecological flow as electrical current, informed by continuous spatial data (a habitat suitability map) (Figure 12). To create the habitat suitability map for London, we created a land use map at 30 m resolution based on a London-wide database of open spaces (Greenspace Information for Greater London CIC 2023c), supplemented by information from OS MasterMap Greenspace Layer, such as the location and extent of private gardens (Ordnance Survey Limited 2022). We assigned resistance values to the different categories of land use, which represent traversal cost, based on the example provided by the Omniscape.jl documentation, our knowledge of land use in Greater London and various trial runs (see Table A6.1 in Appendix 6 for the resistance values). The source strength raster was set as the inverse of the resistance raster, with all open spaces able to act as sources. We set the radius size of the search window to 3 km based on previous studies using Omniscape in an urban environment

(Choe and Tohorne 2019, Park et al. 2023). We did not include the Thames River as a barrier to movement, as it would not be a sufficient barrier for flying animals or to the dispersal of plant seeds. To mitigate for edge effects, we extended the habitat suitability map beyond the boundary of study area (McRae et al 2016). We ran the Omniscape algorithm at a resolution of 30 m, and then averaged the values of connectivity for green spaces (as defined by the Greater London Authority's Green Cover data) within each 300 by 300 m grid.

The above data acquisition steps created a raster stack at 300 m resolution, with each layer representing a different metric. We applied a 300 m interior buffer to the raster, as an edge effect was seen for some of the metrics (especially the landscape metrics) and in some cases the underlying data was not available beyond Greater London, which made extending the study area unfeasible. Finally, we extracted the values from the raster using the function values from the terra package (Hijmans 2022), to create a data frame of metrics for Greater London. Finally, we carried out the data acquisition workflow described above again but using a grid size of 1200 m x 1200 m, and 5000 m x 5000 m. The spatial resolution of 1200 m was chosen as it would be a good choice of resolution for cities where less fine resolution data is available. The spatial resolution of 5000 m was chosen as it is as coarse a scale as is possible whilst allowing for enough grid cells over Greater London for statistical analysis to be meaningful. Both spatial resolutions would be relevant to national scale assessments of assets and benefits (e.g. Maskell et al. 2016, Mancini et al. 2018). We used the same input datasets for the indicators, and followed the same data processing steps, simply changing the size of the grid for their calculation.

4.2.3 Analysis

To test for dependency between the three dimensions of natural capital assets (H1 and H4) we conducted a principal component analysis on the data frame of metrics to understand the correlation structures in the data, using prcomp from core R stats package (R Core Team 2022). We used principal component analysis because it is a good choice to reveal structure in data and relationships when multiple variables are present. We tested for the significance ($p < 0.05$) of the principal components and the loadings using the PCAtest function in the PCAtest package (Camargo 2024), which implements permutation-based statistical tests to evaluate the significance of each PC axis and of contributions of each observed variable to the significant axes (Camargo 2022). We used the function fviz_eig to create a scree plot and the function fviz_pca_var to plot the loadings of the PCA, both from the package factoextra (Kassambara 2020).

We used Pearson correlation to test whether the dependency between quantity and spatial configuration is the strongest (H2) and whether the strength of correlation varies with the indicator chosen (H3), as it allows us to understand the relationship between specific metrics, and we visualised the correlation matrix using the function ggcorrplot in the ggcorrplot package (Kassambara 2023).

4.3 Results

Our results confirm our first hypothesis, as the first principal component of the indicators calculated at 300 m spatial resolution was found to be statistically significant, explaining 56.3% of the variance in the data, and all variables have statistically significant loadings for the first principal component (Table 7), highlighting a strong correlation structure among the variables (Figure 13). The second principal component was also found to be statistically significant, although none of the variables had significant loadings for this axis (Figure A6.1 in Appendix 6).

***Table 7**: Loadings for the first (PC1) and second (PC2) principal components from the indicators for quantity, quality and spatial configuration for Greater London calculated for a 300 m grid, showing the contribution of each variable. The first principal component explains 56.3% of the variance in the data and the second principal component explains* 13.4% *of the variance in the data, and were chosen as they were found to be statistically significant using a permutation-based statistical tests (Camargo 2022). The signs of the loadings are arbitrary, and randomly assigned, so may differ between different programs for principal component analysis (R Core Team 2022).*

Dimension	Indicator	PC1*	PC2*
Quantity	Area green space	-0.36*	0.23
Quality	Species richness	-0.12*	-0.36
Quality	NDVI	-0.26*	-0.36
Quality	Total area of tree canopy cover	-0.29*	-0.35
Quality	Total length of footpaths	-0.24*	-0.32
Quality	Total area of habitat of principal importance	-0.30*	-0.32
Spatial configuration	CAI_AM	-0.36*	0.27
Spatial configuration	LSI	0.33*	-0.32
Spatial configuration	LPI	-0.35*	0.28
Spatial configuration	NP	0.32*	-0.25
Spatial configuration	Landscape connectivity	-0.29*	-0.20

*Indicates statistically significant ($p < 0.05$)

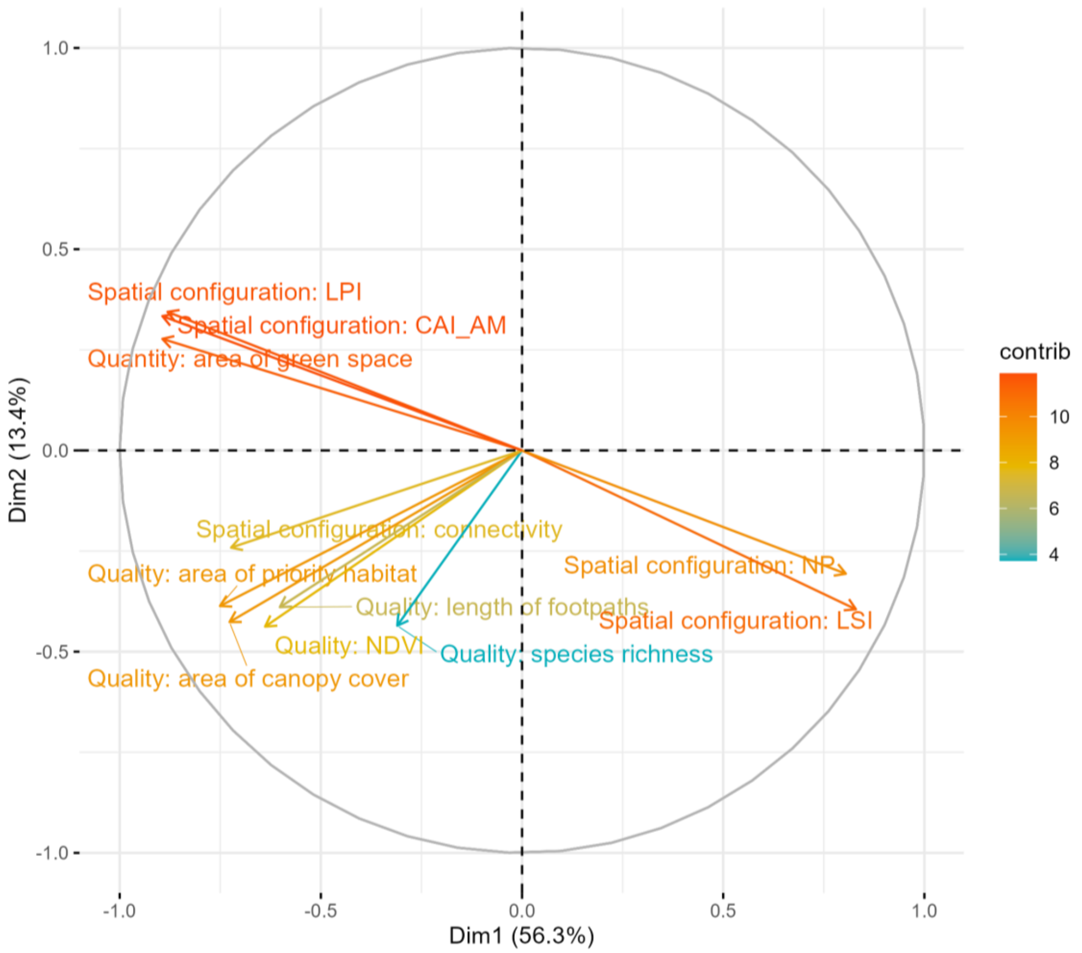


Figure 13: Loadings plot for the principal component analysis of the indicators of quantity, quality, and spatial configuration, calculated at 300 m, showing the contribution of the different variables to the first and second principal component. The angles between the vectors tell us how variables correlate with one another - when two vectors are close, forming a small angle, the two variables they represent are positively correlated. When they diverge and form a large angle (close to 180°), they are negative correlated. The length of the vector indicates the contribution to the principal component.

Our results also broadly support our second hypothesis, with the four highest Pearson correlation coefficients found for the relationship between quantity and spatial configuration (Figure 14), specifically between the total area of green space and the landscape metric CAI_AM ($r = 0.90$, $df = 16812$, $p < 0.001$), the total area of green space and the landscape metric LSI ($r = -0.75$, $df = 16812$, $p < 0.001$), the total area of green space and the landscape metric LPI (r = 0.92, df = 16812, p < 0.001), and the total area of green space and the landscape metric NP (r = -0.76 , df = 16812, p < 0.001). However, the correlation between the total area of greens space and landscape connectivity was not as high (r = 0.52, df = 16812, p < 0.001). All the correlation coefficients were statistically significant (p < 0.001). The fact that a negative correlation was found between total area of green space and LSI and NP is not a problem, as this study is testing for whether indicators are correlated, irrespective of whether the relationship is positive of negative.

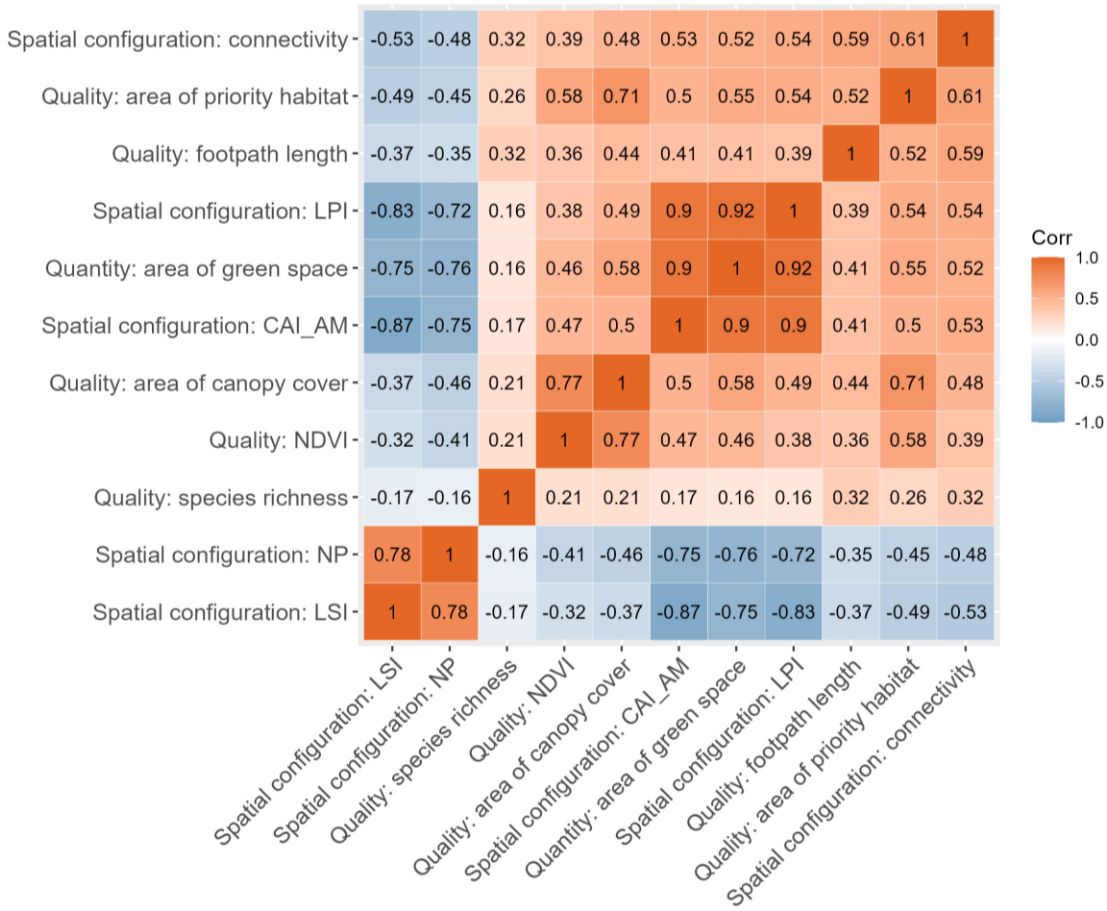


Figure 14: Pearson correlation coefficients for the indicators used in this study, calculated for a 300 m grid. The correlation coefficients were all statistically significant with p < 0.001.

Our results lend support to the third hypothesis, as species richness is only weakly correlated with all the other indicators, including the other indicators for the quality of green space (area of canopy cover and NDVI) (Figure 13, Figure 14). Overall, there is no obvious tendency for bivariate correlations to be higher within pairs of indicators representing a particular dimension, than between pairs of indicators representing different dimensions, with most Pearson correlation coefficients falling within the range of 0.3 and 0.6.

Finally, our results confirm our fourth hypothesis, as the first principal component of the indicators calculated with grid sizes 1200 m and 5000 m were also found to be statistically significant, explaining 60.7% and 54.2% of the variance in the data, respectively, and all of the variables have statistically significant loadings for the first principal component (Appendix 7, Table A7.1), highlighting a strong correlation structure among the variables. The Pearson correlation coefficients between individual indicators is very similar for the indicators calculated using the 1200 m grid to those calculated using the 300 m grid, but this changes at 5000 m spatial resolution, where the indicators for quantity and quality are generally more strongly correlated than for quantity and spatial configuration (Appendix 7, Figure A7.1 a and b).

4.4 Discussion

In this study we assess the strength of dependency between the quantity, quality and spatial configuration of green space in Greater London. Our results show that the indicators for the three dimensions of natural capital have a statistically significant correlation structure, which suggests that it may not be strictly necessary to monitor all three dimensions of natural capital to capture asset condition, and thus resolve how much benefits are affected by deterioration in the condition of natural assets (Mace et al. 2015). However, the strength of dependency varies depending on the choice of indicator for quality and spatial configuration, highlighting that the choice of indicator is an important step and should be informed first and foremost by the purpose of the natural capital assessment or study. Many studies on green spaces in the urban environment are motivated by wanting to understand the benefits they provide, such as recreation or carbon storage, and this should inform which indicators are most relevant. The strength of correlation between the different indicators also changed when they were calculated at a courser resolution (using the 5000 m grid), which is in line with previous work that show that indicators can change with scale in different ways, sometimes predictably and in other cases erratically (Wu et al. 2002). Finally, our work provides further evidence that landscape metrics and the quantity of habitat correlate strongly, and shows that landscape connectivity also correlates with asset quantity, but that this correlation is less strong than for the landscape metrics. The correlation between indicators for the quantity and spatial configuration of natural capital assets poses a challenge for a broader body of work that attempts to partition the explanatory effect of these two dimensions on, for example, communities of carabids (Neumann et al. 2016), the urban heat island effect (Li and Zhou 2019), and landscape susceptibility to processes such as fire or disease (Bierwagen 2007). Some green space urban management plans aim to conserve and expand highly biodiverse areas of the city, to conserve local biodiversity, create opportunities for environmental education, and improve human well-being (Dearborn and Kark 2010). The results from this study show that it is unlikely to be enough to increase the total area of green space within urban areas when the aim is to create more highly biodiverse green spaces, but rather that more active management strategies will be needed, such as reintroducing natural species, planting street verges, fencing, managing threats, tree planting and restoring grasslands (Soanes et al. 2023). However, this may be different in other contexts – for example, in larger rural areas, passive rewilding projects have shown that it is possible for species to reassemble by taking a "wait-and-see" approach (Elphick et al. 2024). It is thus possible that the measures of quantity and species diversity would correlate more strongly in habitats outside of urban areas, as has been suggested elsewhere (Loke et al. 2019)There are various considerations when interpreting the results, especially when drawing implications for management. Importantly, some of the patterns of dependency found in the study are due to historical change with a very different starting point than we have today and are the result of urbanization encroaching onto existing habitat (Foster et al. 2003). Thus, not all the results will be relevant to enhancing the extent, quality and spatial connectivity of existing green spaces today. For example, there are small pocket of high quality remnant habitat in Greater London, but it is likely that these were established in a situation where open green space was more widespread. Whilst ancient woodland or small areas of wetland have survived urban encroachment by receiving protected status, if we want to re-establish new areas like this it is possible that larger areas will be needed to provide enough buffering for species to establish themselves in the face of the pressure of urban edge effects (Wang and Yang 2022), and to mitigate against the added climatic stress brought about by climate change (Weiskopf et al. 2020). In addition to this, the historical land use of urban green areas influences their condition. For example, forest sites in New York City with a history of agriculture, lawn, or built environments were found to have more invasive species groundcover, and the longer a forest had no historical sign of human disturbance, the higher the native basal area (Pregitzer and Bradford 2023). This highlights that what is found when looking at historical change may not always work when implementing future management decisions on preserving, and where possible, reinstating priority natural habitats.

There are also various considerations around the data and indicators used in this study. Firstly, some of the data used to create indicators for quantity and quality – namely the total area of green space, tree canopy area and NDVI – are all sourced from remote sensing images. Whilst they are sourced from different sensors and have undergone different post-processing steps, they still may have underlying similarities in the errors and noise in the data, and some of the correlation between them may be due to them coming from a similar data source. An example of this would be confusing trees and green space in classification, or artificial turf and green space (Crowson et al. 2024). It is likely that choosing to use datasets with different sources is a good way to minimise correlation between indicators. Secondly, a limitation of this study is that the work on connectivity, using the Omniscape.jl algorithm, does not consider the river

Thames to be a barrier for connectivity. This is legitimate for some species, such as birds, that are well represented in our species data. However, for some other species the river Thames represents a real barrier to movement, as would smaller rivers such as the River Lea, and London's canal system. Finally, the species data used to calculate species richness was sourced from records compiled mostly by volunteers, for different organisations and purposed (Greenspace Information for Greater London CIC 2023d). This type of data is known to have a bias due to volunteers' preference for recording particular taxa or at particular sites (Boakes et al. 2016), including a preference towards areas of higher species richness (Tulloch et al. 2013).

To be able to generalise the findings, it would be interesting to see if similar results are found for other cities, with different geographies and histories. For example, Greater London is a comparatively old urban centre, constrained by a green belt. Previous work has shown that older urban areas have had more time for the adverse impacts of urbanisation to be realised (Norton et al. 2016). Younger cities may have larger extinction debts, that is, the number of species expected to go extinct as the adverse impacts of urbanisation are realised over time (Hans et al. 2009), and thus different relationships might be found between the number of species and other indicators. It would also be interesting to see if similar relationships exist between the quantity, quality and spatial configuration for other assets or habitats outside of urban areas, such as forests, where similar questions about quantity, quality and spatial configuration arise. Another interesting avenue of future study would be to attempt to understand the scale effect not only from a change in extent (that is, the different grid sizes used to calculate indicators in this study), but also from a change in zoning (by shifting the grids spatially, without changing their extent) and from the grain size of the underlying data (the effect of raster resolution), as these all contribute to the problems of scale effects and spatial aggregation in landscape ecology (Wu et al. 2002).

4.5 Conclusion

In this study the indicators for the three dimensions of natural capital – quantity, quality and spatial configuration – were shown to be significantly correlated, and thus contain some redundancy. This may mean that it is not useful to implicitly conceptualise the three dimensions of natural capital as mutually independent, as set out in the introduction. Alternatively, the redundancy of information may be because the indicators used in this study are not fully or reliably describing the dimension they set out to capture, due to problems with the underlying data, or other issues around how they are calculated.

Either way, it remains important to use multiple indicators when assessing the condition of urban green spaces, and to choose these based on the purpose of the natural capital assessment or study. The quality, quantity and spatial configuration of natural capital is a useful starting point to consider what characteristics of natural capital assets underpin the benefits they provide. However, other work points to other characteristics of natural capital that are relevant to the benefits we receive, such as pressures on assets (Harrison et al. 2017) and demand for benefits (e.g. recreation, Liu et al. 2020). These aspects are particularly important for urban green spaces, as these areas are placed under intense pressure due to urban development, and are unequally distributed spatially, with high demand and low provision in some neighbourhoods (Wolch et al. 2014, Greenspace Information for Greater London CIC 2023a).

Chapter 5: Discussion and Conclusion

The aim of this thesis was to explore the potential of big data and associated techniques to operationalise the natural capital framework at a national scale in England, through a better understanding of the relationship between natural capital assets and the benefits that flow from them. In Chapter 2, I demonstrated the potential of emerging datasets to capture important aspects of sociocultural value that are otherwise hard to include in a formal valuation process (Figure 15a). In Chapter 3, I showed how a range of different environmental datasets can be combined with statistical techniques to understand the complex ways that assets come together to provide benefits, namely the benefits of good water quality and agricultural production, and how this approach can be used to enable spatially targeted management when these benefits conflict (Figure 15b). Finally, through the work in Chapter 4 I showed that there is significant dependency between the indicators for quantity, quality and spatial configuration of natural capital assets, suggesting that monitoring all three dimensions of natural capital leads to some redundancy of information (Figure 15c).

	(a) Chapter 2	(b) Chapter 3	(c) Chapter 4
Operationalisation of the natural capital framework	• Full valuation of the natural world is challenging, and a range of indicators are needed • This chapter explores a new dataset to capture the sociocultural values of nature at broad spatial scales	• Society receives multiple, sometimes conflicting, benefits from natural capital assets • This chapter uses statistical methods and big data to enable spatially targeted management, to maximise benefit provision in the face of conflicting benefits	• Quantity, quality and spatial configuration of natural capital all need to be monitored to describe asset condition • This chapter explores to what extent the three dimensions of natural capital may be closely related, providing some redundancy in information
Assets and benefits considered	• Asset: designated areas in England • Benefit: Sociocultural value	• Asset: river catchments in England • Benefits: Agricultural food production and good water quality	• Asset: green spaces in Greater London • Benefits: Multiple
Data used	• Location of Flickr photographs • Geotagged Wikipedia pages • Characteristics of designated areas (coastal location, species richness, population density, etc)	• Total Oxidised Nitrogen and Reactive Phosphorous concentrations • Catchment characteristics (land cover maps, agricultural census data, population equivalent of WWTWs in the catchments, etc)	• Maps of green spaces in London • Remote sensing images • Species records • Map of tree canopies
Key aim(s)	• I compare Flickr and Wikipedia data in their ability to characterise the sociocultural value of designated areas in England • I assess how the two datasets relate to species richness.	• I assess population density as a driver of the relative importance of agricultural land use and WWTWs in predicting N and P concentrations nationally	• I assess the strength of dependency between the three dimensions of natural capital (quantity, quality and spatial configuration), using multiple indicators for each dimension
Key findings	• Wikipedia data has the potential to capture important aspects of sociocultural value that are otherwise hard to include in a formal valuation process. • Species richness has a significant positive effect on public interest in designated areas in England	• Statistical methods and big data can be used to enable spatially targeted management when there are conflicts in the provision of multiple benefits • Agricultural sources of N and P dominate in catchments with low population density. • Waste water treatment works dominate in catchments with high population density.	• There is significant dependency between the indicators for quantity, quality and spatial configuration of natural capital assets • Monitoring all three dimensions of natural capital leads to some redundancy of information. • There is potential to maintain highly biodiverse areas in cities, without assigning large areas to this.

Figure 15: Overview of how the three research chapters in this thesis contribute to the operationalisation of the natural capital framework, including the key findings from each chapter.

The most novel finding of the work in Chapter 2, using Wikipedia data to capture the diverse sociocultural values of designated areas in England, is that species richness has a significant positive effect on public interest in designated areas in England, which points to the way people are interested in and have a relationship with species that is not necessarily linked to direct contact with them. The lack of relationship between species richness and visitation is in line with previous research that showed that high natural and conservation value and areas of high recreational value do no tend to overlap (Mancini et al 2019). In terms of visitation, our study adds to an existing body of research on peoples' choices of where they spend their times outdoors, which has also been shown to vary between longer destination trips and more local trips (Graham and Eigenbrod 2019).

The most important finding of the water quality work in Chapter 3 is that population density was shown to be a driver of the relative importance of waste water treatment works and agricultural land use as sources of N and P. Whilst it may seem intuitive that agricultural sources of N and P are likely to dominate in catchments with low population density, compared to point sources from WWTWs, the fact that mean N and P concentrations are lower in these catchments means that WWTWs could still have been comparatively more important. Conversely, whilst previous work has proposed that the agriculture's contribution to N and P concentrations in rivers may be less important than previously thought in densely populated regions (Withers et al. 2014, Foy 2007), I was able to test this for the first time at a national scale. The results feed into a wider debate around the relative contribution of "point sources" of sewage effluent from WWTWs and "diffuse" sources associated with agricultural land use, which is the starting point for environmental management in this area (Neal et al. 2010), although the grouping of anthropogenic sources of N and P into two groups ("point sources" and "diffuse sources") has been accused of being simplistic in a number of ways (Withers and Jarvie 2008). Other drivers of context dependency are likely to exist, of course, and we could have split our data into groups in various ways – for example, watersheds with high or low precipitation, and watersheds with geology that is more or less permeable. I carried out initial investigations into the effect of forest land cover on N and P concentration, and the effect was not strong enough to be picked up at a national level by the models, which shows that some drivers of context dependency are stronger than others at a landscape scale (Spake et al. 2019).

Finally, the work in Chapter 4 on the quantity, quality and spatial configuration of green space in Greater London data is novel in terms of its focus on the urban environment, which is often overlooked within studies of biodiversity and natural capital. This also means that some of the findings may be unique to urban environments. The most unexpected finding from this work was the way that the indicator derived from species data did not correlate with other indicators, including the quantity (area) of green space, whilst reliably picking out high biodiversity area in London (based on visual inspection of the species richness data and my own knowledge of green spaces in London). This shows that there is potential to maintain highly biodiverse areas in cities, without needing to assign large areas for this. However, other benefits will scale more linearly with the quantity and primary productivity of green space in the city, such as the cooling effect of green spaces (Kong et al. 2014).

5.1 Lessons learnt and recommendations.

A common theme emerging from the research presented here is that the purpose of the natural capital assessment should inform how it is designed and carried out, including the choice of values we wish to capture, the characteristics of assets we are most interested in, the data sources and indicators that are best to use, and how robust the validation of results needs to be. As was highlighted in the introduction, natural capital assessments vary in their purpose within science and policy, including: regular monitoring exercises to document change (Office for National Statistics 2023a), accounting exercises that aim to negotiate specific trade-offs (Posner et al. 2016, Chan et al. 2020), valuation exercised to enhance the visibility and communication of cultural ecosystem services (Hernández-Morcillo et al. 2013, Hinson et al. 2022), and studies to better understand the underlying processes of how assets provide benefits to society (e.g. Spake et al. 2019). This both explains and justifies the wide range of approaches, data sources and indicators that have been used within the natural capital approach so far (Fairbrass et al. 2020).

Whilst many of the decisions around the design of a natural capital assessment will be specific to its objectives, there are some generalisations that can be made based on the findings presented in this thesis. Firstly, whilst the work on the indicators of quantity, quality and spatial configuration suggests that it may not be necessary to monitor all dimensions in all cases, as I found some redundancy of information between the indicators for each dimension, taken together our work strongly suggests that multiple indicators are necessary to capture the value of nature. This is a way to mitigate the limitations of individual indicators, and it means the valuation exercise will capture different aspects of asset status and benefits, providing a fuller picture. This is already the approach taken within biodiversity credits (rePLANET 2024) and is similar to the way that consumer inflation is measured using the consumer price index, which is calculated by selecting household consumed goods and services that are typically purchased by specific groups of households, grouping them into a "basket", and periodically tracking percentage changes in the cost of buying that "basket" (OECD 2023). The choice of goods and services included in the "basket" will vary from country to country, but that does not matter, as it is what people are buying in that place.

Secondly, when choosing different environmental datasets to use within a natural capital assessment or study, I would recommend prioritising datasets with different collection methods or sources whenever possible, as this will make it less likely that they contain similar noise due to the collection and processing method. Whilst certain data types, such as remote sensing data, are increasingly available and accessible, it is important to continue to explore other sources of data, and continue to fund fieldwork and other monitoring programmes to ensure diverse datasets are available for indicator creation, as well as to ground truth remote sensing data. Finally, many datasets, including widely used data such as land cover maps, are datasets derived from a model, rather than raw observations. Some models are better than others at using what is measured to create datasets and indicators (which can be thought of as "data quality" or "measurement error"), but there is also a risk of "concept error" if we are uncritical about how we use indicators to represent the concept of interest. Thus, indicators need to be interpreted with reference to the reason they were chosen and their original data source, as these things will determine what they are likely to be able to capture and not. Understanding and communicating this nuance is also a way to reduce the risk of the natural capital approach being used against conservation efforts, by arguing, for example, that the absence of a certain indicators implies an absence of value, or by monetising benefits in a way that is not appropriate.

5.2 Limitations of the approach

A limitation of the work done for this thesis is that not much consideration has been given to the temporal dimension of natural capital monitoring. Whilst the exact purpose of natural capital assessments varies, there is a common theme running through them of wanting to safeguard the benefits we receive from nature for future generations. To assess progress towards this goal it is necessary to be able to detect change and identify trends. For example, can changes to the sociocultural values of designated areas be detected by new and emerging datasets such as Wikipedia? We use the location of Wikipedia pages as our data source in Chapter 2, and to introduce a temporal dimension it would probably be necessary to study the number of page views of a subset of Wikipedia pages.

In terms of the water quality work in Chapter 3, the next step would be to use the statistical models to predict the change in N and P concentrations that would be brought about by a specific land use change scenario or where known changes in land use or WWTW treatment have occurred (e.g. Civan et al. 2018), which would allow us to assess whether these changes are likely to be detectable and distinguishable from temporary fluctuations. Climate change is also an important consideration in this context, firstly because the effects of climate change need to be separated from those resulting from management decisions. Secondly, climate change may shift the importance of particular sources of N and P from those found in this study. Initial research in this area shows that the two sources are affected by climate change in different ways, with a stronger effect of climate change on nutrient concentration in catchments where point sources dominate, as reduced flows lead to less dilution of sewage inputs (Wade et al. 2022). This may further strengthen the importance of WWTWs in catchments with high population density.

In terms of Chapter 4, it would be interesting to assess which of the indicators for the quantity, quality and spatial configuration of green space allow change detection, and thus detect progress (or lack of progress) towards policy targets and detect the effects of climate change. Increases in drought due to climate change will likely have a strong effect on NDVI, as NDVI can detect changes to the health and composition of vegetation in green spaces (Gascon et al. 2016). Species diversity might also change with climate change, particularly due to invasive species (Sparks et al. 2007, Lockwood et al. 2009). However, as an indicator species diversity will still be able to pick out high biodiversity areas. To detect new species present in green spaces or an increase in abundance of certain species it is likely that the best approach would be to harness the potential of crowdsourced data through species identification apps and other citizen science initiatives (Isaac et al. 2014), and consider a wider range of species.

Being able to reliably detect change through time is a high bar against which to judge data and indicators. Detecting trend in species data is challenging (Isaac et al. 2014, Kéry et al. 2009), and so is accurately detecting land cover change in England using remote sensing data (Marston et al. 2023), despite land cover mapping being an area that receives a lot of resources. This is because accurately detecting change between classified habitat or land cover maps is not as straightforward as comparing maps from different time points, as real change needs to be differentiated from those arising due to errors and variation in methodology, and thus the error rate of the change detection map is larger than that of the individual maps being compared (Barber and Robinson 2023). Land cover maps also do not give information on land use, so changes to agricultural intensity or habitat quality are not identified when comparing land cover maps. Given the difficulty in detecting change and trends using even established data types and methods, it is perhaps unsurprising that the work in this thesis exploring new opportunities presented by data on the environment has not included progress on change detection, and future work specifically on the temporal aspect of natural capital assessment is much needed.

Another limitation of this theses is that it focuses only on England, which is not necessarily representative of other parts of the world. Many of the data sources used in this thesis would not be as widely available in other countries, including data from the photo sharing platform Flickr, and much of the data used in the study of green spaces in London. The rate of change in asset extent and condition is also much smaller in London than it is in the Amazon rainforest, for example, so some of the difficulties of detecting change in England is specific to a country where land cover change and changes to species composition are currently only happening comparatively slowly. This means that whilst this thesis contributes to the operationalisation of the natural capital approach in England, the approach and questions would be different for work supporting international environmental accounting efforts.

5.3 Technical challenges to taking a big data approach within natural capital assessments

Data on the natural environment in England are comparatively abundant and accessible, but there is still a limit to data available from long term monitoring programmes, particularly at a national scale. Even when an abundance of data seems to be available in the first instance, once various criteria have been applied to achieve a minimum data standard and assure that measurements are independent of each other, datasets can shrink considerably, as we found was the case for monitoring stations from the Water Quality Archive (Environment Agency 2021) in Chapter 3. This highlights the continued technical challenges involved in attempting to apply machine learning methods in the environmental sciences, where data are costly to collect, and funding is limited. The availability of good quality, representative data becomes an ethical consideration when indicators and models go on to inform management decisions, as they need to undergo extensive independent validation to be trusted. As described in the introduction, big data broadly refers to the increasing volume, variety, and velocity of data streams over the past 20 years or so (Hampton et al. 2013, Chen et al. 2014), and I would say that the work presented in this thesis has focused especially on the variety of environmental data, looking for ways to combine different datasets to tackle scientific questions and operationalise the natural capital framework within policy, as well as exploring new sources of data to add to this variety. To apply big data techniques that make use of very large volumes of data and harness the velocity at which is it possible to capture data, such as deep learning techniques based on timeseries analysis (Cheng et al. 2023) or sensors with onboard processing capabilities (Mahendra et al. 2020), it is necessary to identify areas that are likely to benefit from these kind of techniques and carry out targeted data collection.

Another technical obstacle to applying some of the methods used in this thesis more broadly is the time needed to do the data processing and preparation, which would be difficult to justify in a more applied setting. Our studies involved a considerable effort in terms of data cleaning and preparation. There is potential for future developments in large language model-based chatbots and other artificial intelligent tools to speed up some of the data processing involved (e.g. Agathokleous et al. 2023, Hassani et al. 2023). This, along with further availability of large datasets, may facilitate a wider uptake of the type of methods explored in this thesis in the future. There is also the option of including methods and data into decision making tools or packages to make them more widely available. Flickr data can now be accessed through the decision-making tool InVEST (Sharp et al. 2018) and the photosearcher R package (Fox et al. 2020), for example.

Finally, our work using statistical techniques, including machine learning techniques, within a natural capital approach has shown that there are still challenges to overcome when using them. Statistical techniques need not stand alone, however, and can be used alongside more traditional models - as data exploration, validation, or as a component in a larger model framework (or digital twin) (Blair and Henrys 2023). Existing mixed approaches in other domains of environmental science and ecology include using more traditional process-based models to build features or indicators to be used in statistical models (e.g. Mavromatis 2014), and using statistical approaches to estimate unknown parameters for process-based models (Maestrini et al. 2022), particularly in situations where there is limited capacity to describe the process in physical equations (Saha et al. 2021). Moving forward, it is important to recognise that new data and techniques will not necessarily replace previous approaches but complement them.

5.4. What's next?

There are various new developments brought about by the big data era that are starting to be applied within ecology (McCrea et al. 2022) and conservation (Runting et al. 2020), and could potentially support the natural capital approach. An example of this is the emergence of deep learning, which is likely to have an important role in the future direction of methods in ecology and beyond (Pichler and Hartig, 2022, Perry et al. 2022, Havinga et al. 2023). Deep learning is as a family of machine learning algorithms that are composed of multiple processing layers that transfer input to output by progressively learning higher level features. Because of the ability of deep learning to automatically discovering the most important data features and relevant patterns (Borowiec et al. 2022), some anticipate it will revolutionise how remote sensing data can be used (Li et al. 2017), and deep learning approaches are already being used to fill knowledge gaps using camara traps, bioacoustics and social media data (see e.g. Willi et al. 2018, Stowell 2022, Havinga et al. 2023). To overcome the technical challenge of having to provide very large datasets to train, test and validate deep learning algorithms, there is work being carried out to develop foundation models (e.g. Stewart et al. 2023 for remote sensing data), inspired by the capabilities demonstrated by foundation models in other realms, including GPT (which underpins the chatbot ChatGPT) and Dall-E (which creates images from text). Foundation models are pre-trained to perform a general task (e.g. land cover classification) and can then be fine-tuned for specific applications, although these are still at an early stage of development. Whilst these developments are exciting, there is still a need to identify promising avenues for their implementation (Pettorelli et al. in press).

5.5 Conclusion

The natural capital approach remains central to environmental policy and discourse in England, and environmental accounting is still in use internationally, with a shift in awareness towards the diverse values of nature (IPBES 2022). The opportunities provided by big data have been shown in this thesis to expand the diversity of values that can be included in the valuation process, increase our understanding of the relationship between assets and the benefits we receive from them, and help inform monitoring and management efforts in multiple ways. However, operationalising the natural capital approach remains challenging, particularly with regards to tracking changes in assets and their status over time in a meaningful way, and with regards to the need for data to validate models and develop methods. There remains a considerable gap between the ambitions and outcomes of national policy on the environment (Natural Capital Committee 2019). Whilst information is important to inform policy and decision making, accounting exercises need to be backed by legislation, public support and compliance monitoring if we want to see the large-scale changes to environmental stewardship required to be the first generation to leave the environment in a better state than we received it, as was lain out in the 25 Year Environment Plan for England (Defra 2018a).

References

Adams, W. M., Hodge, I. D., and Sandbrook, L., 2014. New spaces for nature : the re-territorialisation of biodiversity conservation under neoliberalism in the UK. Transactions of the Institute of British Geographers, 39: 574-588.

Agathokleous, E., Saitanis, C. J., Fang, C., and Yu, Z., 2023. Use of ChatGPT: What does it mean for biology and environmental science? Science of the Total Environment, 888(May), 164154. https://doi.org/10.1016/j.scitotenv.2023.164154

Allam, K., and Naser, H., 2018. Synergy Among Sustainable Development Goals and Biodiversity-Related Conventions in Bahrain. Retrieved from https://portals.iucn.org/library/sites/library/files/documents/2018-016-En.pdf

Allen, K. E., Castellano, C., and Pessagno, S., 2021. Using dialogue to contextualize culture, ecosystem services, and cultural ecosystem services. Ecology and Society, 26(2). https://doi.org/10.5751/ES-12187-260207

Andrew, M. E., Wulder, M. A., and Nelson, T. A., 2014. Potential contributions of remote sensing to ecosystem service assessments. Progress in Physical Geography, 38(3), 328–353. https://doi.org/10.1177/0309133314528942

Angus, J., Baldock, J., Farrel, M., and Norton, R., 2013. Plant-Available Nitrogen Fact Sheet Southern Region. Retrieved from: www.grdc.com.au/GRDC-FS-PlantAvailableNitrogen

ARIES, 2019. Artificial Intelligence for Ecosystem Services. Retrieved from http://aries.integratedmodelling.org/

Ariluoma, M., Ottelin, J., Hautamäki, R., Tuhkanen, E. M., and Mänttäri, M., 2021. Carbon sequestration and storage potential of urban green in residential yards: A case study from Helsinki. Urban Forestry and Urban Greening, 57. https://doi.org/10.1016/j.ufug.2020.126939

Armitage, D. W., and Ober, H. K., 2010. A comparison of supervised learning techniques in the classification of bat echolocation calls. Ecological Informatics, 5(6), 465–473. https://doi.org/10.1016/j.ecoinf.2010.08.001

Ayanu, Y. Z., Conrad, C., Nauss, T., Wegmann, M., and Koellner, T., 2012. Quantifying and mapping ecosystem services supplies and demands: A review of remote sensing applications. Environmental Science and Technology, 46(16), 8529–8541. https://doi.org/10.1021/es300157u

Balmford, A., Beresford, J., Green, J., Naidoo, R., Walpole, M., and Manica, A., 2009. A global perspective on trends in nature-based tourism. PLoS Biology, 7(6), 1–6. https://doi.org/10.1371/journal.pbio.1000144

Barber, R., and Robinson, P., 2023. Natural Capital Ecosystem Assessment: NE1.6 Change Detection Scoping object-based change detection for Living. (March), 1–52. Retrieved from: https://publications.naturalengland.org.uk/publication/6002530300198912

Barbier, E. B., 2013. Wealth accounting, ecological capital and ecosystem services. Environment and Development Economics, 18(2), 133–161. https://doi.org/10.1017/S1355770X12000551

Bayliss, A. C., 1999. Catchment descriptors. Volume 5 of the Flood Estimation Handbook. Centre for Ecology & Hydrology.

Belgiu, M., and Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS Journal of Photogrammetry and Remote Sensing, 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bell, V. A., Naden, P. S., Tipping, E., Davies, H. N., Carnell, E., Davies, J. A. C. et al., 2021. Long term simulations of macronutrients (C, N and P) in UK freshwaters. Science of the Total Environment, 776. https://doi.org/10.1016/j.scitotenv.2021.145813

Bierwagen, B. G., 2007. Connectivity in urbanizing landscapes: The importance of habitat configuration, urban area size, and dispersal. Urban Ecosystems, 10(1), 29–42. https://doi.org/10.1007/s11252-006-0011-6

Birmingham City Council, 2013. Green Living Spaces Plan. (September 2013), 31. Retrieved from https://www.birmingham.gov.uk/greenlivingspaces

Blair, G. S., 2021. Digital twins of the natural environment. Patterns, 2(10), 100359. https://doi.org/10.1016/j.patter.2021.100359

Blair, G. S., and Henrys, P. A., 2023. The role of data science in environmental digital twins: In praise of the arrows. Environmetrics, 34(2), 1–6. https://doi.org/10.1002/env.2789

Boakes, E. H., Gliozzo, G., Seymour, V., Harvey, M., Smith, C., Roy, D. B., and Haklay, M., 2016. Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. Scientific Reports, 6(April), 1–11. https://doi.org/10.1038/srep33051

Boorman, D., Hollis, J. M. and Lilly, A., 1995. Hydrology of Soil Types: A Hydrologically-Based Classification of the Soils of United Kingdom. Institute of Hydrology, Wallingford. doi:10.1029/ 98GL02804.

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., and White, A. E., 2022. Deep learning as a tool for ecology and evolution. Methods in Ecology and Evolution, 13(8), 1640–1660. https://doi.org/10.1111/2041-210X.13901

Braat, L. C., 2018. Five reasons why the Science publication "Assessing nature's contributions to people" (Diaz et al. 2018) would not have been accepted in Ecosystem Services. Ecosystem Services, 30, A1–A2. https://doi.org/https://doi.org/10.1016/j.ecoser.2018.02.002

Breadboard Labs, 2018. Curio Canopy - London Tree Canopy Cover. Retrieved from: https://data.london.gov.uk/dataset/curio-canopy?_gl=1%2a1g83f44%2a_ga%2aMzI3MzA0NzE1LjE2NzU5Mzk0ODI.

Breiman, L., 2001. Statistical modeling: The two cultures. Statistical Science, 16(3), 199–215. https://doi.org/10.1214/ss/1009213726

Bright, G., Connors, E., and Grice, J., 2019. Measuring natural capital: towards accounts for the UK and a basis for improved decision-making. https://doi.org/10.1093/oxrep/gry022

Britannica, 2023. Climate of London, Retrieved 7 February from: https://www.britannica.com/place/London/Climate

Browning, E., Gib, R., Glover-Kapfer, P., and Jones, K. E., 2017. Passive acoustic monitoring in ecology and conservation. WWF Conservation Technology Series, 1(2), 75. https://doi.org/10.13140/RG.2.2.18158.46409

Calcagni, F., Terra, A., Maia, A., John, J., Connolly, T., and Langemeyer, J., 2019. Digital co - construction of relational values : understanding the role of social media for sustainability. Sustainability Science, 14(5), 1309–1321.

Camargo A., 2022. PCAtest: testing the statistical significance of Principal Component Analysis in R. PeerJ, 10:e12967. https://doi.org/10.7717/peerj.12967

Camargo, A., 2024. PCAtest: Statistical significance of PCA. R package version 0.0.1. https://github.com/arleyc/PCAtest

Campbell, G., Bradshaw, M., Brisland, L., Green, C. M., Leach, T. et al., 2020. Reimagining rainwater in parks and green spaces Business in the Community. https://www.london.gov.uk/sites/default/files/gla_office_suds_sector_guide_v3.pdf

Carpenter, S. R., Stanley, E. H., and Vander Zanden, M. J., 2011. State of the world's freshwater ecosystems: Physical, chemical, and biological changes. Annual Review of Environment and Resources, 36, 75–99. https://doi.org/10.1146/annurev-environ-021810-094524

Carpio-Pinedo, J., 2014. Urban Bus Demand Forecast at Stop Level: Space Syntax and Other Built Environment Factors. Evidence from Madrid. Procedia - Social and Behavioral Sciences, 160, 205–214. https://doi.org/10.1016/j.sbspro.2014.12.132

CEH and Natural England, 2017. Mapping Natural Capital (RP02404). Retrieved June 30, 2019, from https://eip.ceh.ac.uk/naturalengland-ncmaps

Chan, K. M. A., and Satterfield, T., 2020. The maturation of ecosystem services: Social and policy research expands, but whither biophysically informed valuation? People and Nature, 2(4), 1021–1060. https://doi.org/10.1002/pan3.10137

Chen, M., Mao, S., and Liu, Y., 2014. Big data: A survey. Mobile Networks and Applications, 19(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chen, A., Yao, L., Sun, R., and Chen, L., 2014. How many metrics are required to identify the effects of the landscape pattern on land surface temperature? Ecological Indicators, 45, 424–433. https://doi.org/10.1016/j.ecolind.2014.05.002

Cheng, X., Sun, Y., Zhang, W., Wang, Y., Cao, X., and Wang, Y.,2023. Application of Deep Learning in Multitemporal Remote Sensing Image Classification. In *Remote Sensing* (Vol. 15). https://doi.org/10.3390/rs15153859

Cheng, X., Van Damme, S., Li, L., and Uyttenhove, P., 2019. Evaluation of cultural ecosystem services: A review of methods. Ecosystem Services, 37(June 2019), 100925.

Choe, H., and Thorne, J. H., 2019. Omnidirectional connectivity of urban open spaces provides context for local government redevelopment plans. Landscape and Ecological Engineering, 15(3), 245–251. https://doi.org/10.1007/s11355-019-00377-8

Chua, M. A. H., Tan, A., and Carrasco, L. R., 2021. Species awareness days: Do people care or are we preaching to the choir? Biological Conservation, 255(March), 109002.

Church, A., Burgess, J., and Ravenscroft, N., 2011. Cultural services. UK national ecosystem assessment. https://doi.org/10.4135/9781483346663.n147

CONABIO, 2017. Sintesis Capital Natural de Mexico: Evaluacion del conocimiento y tendencias de cambio, perspectivas de sustentabilidad, capacidades humanas e institucionales. Retrieved from https://www.biodiversidad.gob.mx/pais/capitalNatMex.html

Copernicus Data Space Ecosystem, 2023. Retrieved February 6, 2024 from: https://dataspace.copernicus.eu/analyse

Copernicus Research and User Support, 2024. Copernicus Research and User Support (RUS) Service portal. Retrieved February 6, 2024, from https://land.copernicus.eu/global/virtual-machine-desktop

Correia, R. A., Ladle, R., Jarić, I., Malhado, A. C. M., Mittermeier, J. C., Roll, U., Soriano-Redondo, A., Veríssimo, D., Fink, C., Hausmann, A. et al, 2021. Digital data sources and methods for conservation culturomics. Conservation Biology, 35(2), 398–411. https://doi.org/10.1111/cobi.13706

Costanza, R., Quatrini, S., and Øystese, S., 2012. Response to George Monbiot: The valuation of nature and ecosystem services is not privatization. Retrieved June 30, 2019, from https://www.climatechangenews.com/2012/08/15/response-to-monbiot-valuation-is-not-privatization/

Countryside Survey, 2019. Retrieved July 1, 2019, from https://countrysidesurvey.org.uk

Cutler, D. R., Edwards Jr., T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J., 2007. Random Forest for Classification in Ecology, Ecology, 88(11), 2783–2792. https://doi.org/https://doi.org/10.1890/07-0539.1

Dasgupta, P., 2021, The Economics of Biodiversity: The Dasgupta Review. Abridged Version. (London: HM Treasury). Retrieved from: https://assets.publishing.service.gov.uk/media/602e92b2e90e07660f807b47/The_Economics_of_Biodiversity_The_Dasgupta_Review_Full_Report.pdf

Davies, H., and Neal, C., 2004. GIS-based methodologies for assessing nitrate, nitrite and ammonium distributions across a major UK basin, the Humber. Hydrology and Earth System Sciences, 8(4), 823–833. https://doi.org/10.5194/hess-8-823-2004

Davies, H., and Neal, C., 2007. Estimating nutrient concentrations from catchment characteristics across the UK. Hydrology and Earth System Sciences, 11(1), 550–558. https://doi.org/10.5194/hess-11-550-2007

Dearborn, D. C., and Kark, S., 2010. Motivaciones para conservar la biodiversidad urbana. Conservation Biology, 24(2), 432–440. https://doi.org/10.1111/j.1523-1739.2009.01328.x

de Bell, S., White, M., Griffiths, A., Darlow, A., Taylor, T., Wheeler, B., and Lovell, R., 2020. Spending time in the garden is positively associated with health and wellbeing: Results from a national survey in England. Landscape and Urban Planning, 200 (June 2019). https://doi.org/10.1016/j.landurbplan.2020.103836

Defra, 2011. The Natural Choice: securing the value of nature. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/228842/8082.pdf

Defra, 2013. Payments for Ecosystem Services: A Best Practice Guide. https://doi.org/10.2305/IUCN.CH.2005.PAG.13.en

Defra, 2014. Water Framework Directive implementation in England and Wales: new and updated standards to protect the water environment. In Department for Environment Food and Rural Affairs. Retrieved from: https://assets.publishing.service.gov.uk/media/5a7f00e740f0b62305b84861/river-basin-planning-standards.pdf

Defra, 2018a. 25-year-environment-plan. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/693158/25-year-environment-plan.pdf

Defra, 2018b. Measuring environmental change - draft indicators framework for the 25 Year Environment Plan Draft for discussion. (December). Retrieved from: https://assets.publishing.service.gov.uk/media/5f6da2108fa8f541eff195a2/indicator-framework-consult-document.pdf

Defra, 2024a. Chapter 7: Crops, Agriculture in the United Kingdom, https://www.gov.uk/government/statistics/agriculture-in-the-united-kingdom-2021/chapter-7-crops

Defra, 2024b. Chapter 8: Livestock, https://www.gov.uk/government/statistics/agriculture-in-the-united-kingdom-2021/chapter-8-livestock

Defra Catchment Explorer, 2023. https://environment.data.gov.uk/catchment-planning/

de Groot, R., Costanza, R., Braat, L., Brander, L., Burkhard, B., Carrasco, L., van Weelden, Louise Willemen, M., 2018. RE: Ecosystem Services are Nature's Contributions to People. Retrieved from Science Letter response website: https://science.sciencemag.org/content/359/6373/270/tab-e-letters

Díaz, S., Demissew, S., Carabias, J., Joly, C., Lonsdale, M., Ash, N., Larigauderie, A., Adhikari, J. R., Arico, S., Báldi, A. et al, 2015. The IPBES Conceptual Framework - connecting nature and people. Current Opinion in Environmental Sustainability, 14, 1–16. https://doi.org/10.1016/j.cosust.2014.11.002

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R. T., Molnár, Z. et al., 2018. Assessing nature's contributions to people. Science, 359(6373), 270–272. https://doi.org/10.1126/science.aap8826

Dickie, I., Cryle, P., Maskell, L., Jones, A. G., Mcvittie, A., Luisetti, T. et al., 2014. UK NEAFO Work Package. Retrieved from http://uknea.unep-wcmc.org/LinkClick.aspx?fileticket=ALFqJld0K8o%3D&tabid=82

Dickson, B., Blaney, R. Miles, L., Regan, E., van Soesbergen, A., Väänänen, E., Blyth, S., Harfoot, M. Martin, C.S., McOwen, C., Newbold, T., van Bochove, J., 2014. Towards a global map of natural capital: key ecosystem assets. Retrieved from http://wcmc.io/towards_a_global_map_of_natural_capital

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., and Kulkarni, R., 2021. How to address data privacy concerns when using social media data in conservation science. Conservation Biology, 35(2), 437–446. https://doi.org/10.1111/cobi.13708

Do, Y., 2019. Valuating aesthetic benefits of cultural ecosystem services using conservation culturomics. Ecosystem Services, 36(April) 100894.

du Toit, J. T., and Pettorelli, N., 2019. The differences between rewilding and restoring an ecologically degraded landscape. Journal of Applied Ecology, 56(11), 2467–2471. https://doi.org/10.1111/1365-2664.13487

Dyer, R., and Oliver, T., 2016. UK ecological status map version 2. Retrieved from https://doi.org/10.5285/58b248a8-6e34-4ffb-ae32-3744566399a2

Ecosystems Knowledge Network, E. K., 2019. Tool Assessor. Retrieved June 30, 2019, from https://ecosystemsknowledge.net/tool

eftec, 2012. Scoping Study To Develop Understanding of a Natural Capital Asset Check. Retrieved from: https://nora.nerc.ac.uk/id/eprint/21296/1/N021296CR.pdf

eftec, 2017. A study to scope and develop urban natural capital accounts for the UK. Retrieved from: https://www.scribd.com/document/737719057/a-Study-to-Scope-and-Develop-Urban-Natural-Capital-Accounts

Elphick, A., Ockendon, N., Aliácar, S., Crowson, M., and Pettorelli, N., 2024. Long-term vegetation trajectories to inform nature recovery strategies: The Greater Côa Valley as a case study. Journal of Environmental Management, 355(October 2023). https://doi.org/10.1016/j.jenvman.2024.120413

Environment Act, 2021, c. 30. Available at: https://www.legislation.gov.uk/ukpga/2021/30/contents (Accessed: 01 January 2024).

Environment Agency, 2019. Phosphorus and Freshwater Eutrophication Pressure Narrative. Retrieved from https://consult.environment-agency.gov.uk/environment-and-business/challenges-and-choices/user_uploads/phosphorus-pressure-rbmp-2021.pdf

Environment Agency, 2021. Environment Agency water quality data from the Water Quality Data Archive. Accessed at: https://environment.data.gov.uk/water-quality/view/landing

Environment Agency, 2023a. Natural Capital Register and Account Tool: summary. Retrieved from: https://www.gov.uk/government/publications/natural-capital-register-and-account-tool/natural-capital-register-and-account-tool-summary

Environment Agency, 2023b. Urban Waste Water Treatment Directive Treatment Plants dataset, available at: https://www.data.gov.uk/dataset/0f76a1c3-1368-476b-a4df-7ef32bfd9a8b/urban-waste-water-treatment-directive-treatment-plants

England Agricultural Census 2016 5km [SHAPE geospatial data], Scale 1:10000, Tiles: GB, Updated: 30 June 2016, Defra, Using: EDINA Agcensus Digimap Service, <https://digimap.edina.ac.uk>, Downloaded: 2022-12-14 14:01:24.586

ESRI, 2022. ArcGIS Pro 3.1.0. Redlands, CA: Environmental Systems Research Institute.

European Commission, 2016. Priority Substances and Certain Other Pollutants according to Annex II of Directive 2008/105/EC. Retrieved June 30, 2019, from http://ec.europa.eu/environment/water/water-framework/priority_substances.htm

European Parliament and Council, 2013. Decision no.1386/2013/EU of the European Parliament and the Council of 20 November 2013 on a General Union Environment Action Programme to 2020. "Living well, within the limits of our planet." Office of the European Union, (1600), 87. https://doi.org/10.2779/66315

Fairbrass, A., Mace, G., Ekins, P., and Milligan, B., 2020. The natural capital indicator framework (NCIF) for improved national natural capital reporting. Ecosystem Services, 46(November 2020), 101198. https://doi.org/10.1016/j.ecoser.2020.101198

FAO, 2019. The State of the World's Biodiversity for Food and Agriculture. In J. B. and D. Pilling (Ed.), FAO Commission on Genetic Resources for Food and Agriculture Assessments (p. 572). Retrieved from http://www.fao.org/3/CA3129EN/CA3129EN.pdf

Farley, J., and Costanza, R., 2010. Payments for ecosystem services: From local to global. Ecological Economics, 69(11), 2060–2068. https://doi.org/10.1016/j.ecolecon.2010.06.010

Forest Research, 2024. Habitat fragmentation – Practical considerations. Retrieved from: https://www.forestresearch.gov.uk/tools-and-resources/fthr/urban-regeneration-and-greenspace-partnership/greenspace-in-practice/practical-considerations-and-challenges-to-greenspace/habitat-fragmentation-practical-considerations/

Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D., and Knapp, A., 2003. The importance of land-use legacies to ecology and conservation. BioScience, 53(1), 77–88. https://doi.org/10.1641/0006-3568(2003)053[0077:TIOLUL]2.0.CO;2

Foy, R. H., 2007. Variation in the reactive phosphorus concentrations in rivers of northwest Europe with respect to their potential to cause eutrophication. Soil Use and Management, 23(SUPPL. 1), 195–204. https://doi.org/10.1111/j.1475-2743.2007.00111.x

Fox, N., August, T., Mancini, F., Parks, K. E., Eigenbrod, F., Bullock, J. M., Sutter, L., Graham, L. J., 2020. "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr. SoftwareX, 12, 100624. https://doi.org/10.1016/j.softx.2020.10062

García-Antúnez, O., Lindgaard, J., Lampinen, J., and Stahl Olafsson, A., 2023. Gardening for wildlife: A mixed-methods exploration of the factors underlying engagement in wildlife-friendly gardening. People and Nature, 5, 808–825. https://doi.org/10.1002/pan3.10450

Gascon, M., Cirach, M., Martínez, D., Dadvand, P., Valentín, A., Plasència, A., and Nieuwenhuijsen, M. J., 2016. Normalized difference vegetation index (NDVI) as a marker of surrounding greenness in epidemiological studies: The case of Barcelona city. Urban Forestry and Urban Greening, 19, 88–94. https://doi.org/10.1016/j.ufug.2016.07.001

Ghermandi, A., and Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: A systematic map. Global Environmental Change, 55, 36-47. https://doi.org/10.1016/j.gloenvcha.2019.02.003

GLA City Intelligence, 2019. London Green and Blue Cover dataset. Retrieved December 2023 from: https://data.london.gov.uk/dataset/green-and-blue-cover

GLA Data Store, 2023. Greater London Data Store. Retrieved December 2023 from: https://data.london.gov.uk/dataset/londons-population

Gliozzo, G., Pettorelli, N., and Muki Haklay, M., 2016. Using crowdsourced imagery to detect cultural ecosystem services: A case study in South Wales, UK. Ecology and Society, 21(3). https://doi.org/10.5751/ES-08436-210306

Gómez-Baggethun, E., and Barton, D. N., 2013. Classifying and valuing ecosystem services for urban planning. Ecological Economics, 86, 235–245. https://doi.org/10.1016/j.ecolecon.2012.08.019

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Goulder, L. H., and Kennedy, D., 1997. Valuing ecosystem services: philosophical bases and empirical methods. In G. C. Daily (Ed.), Nature's services: societal dependence on natural ecosystems (pp. 23–47). Washington, DC: Island Press.

Grafius D.R., Corstanje, R. and Harris, J.A., 2018. Linking ecosystem services, urban form and green space configuration using multivariate landscape metric analysis. Landscape Ecology 33:1–17

Graham, L. J., and Eigenbrod, F., 2019. Scale dependency in drivers of outdoor recreation in England. People and Nature, 1(3), 406–416. https://doi.org/10.1002/pan3.10042

GRASS Development Team, 2022. Geographic Resources Analysis Support System (GRASS) Software, Version 8.2.0. Open Source Geospatial Foundation. http://grass.osgeo.org

Graves, R. A., Pearson, S. M., and Turner, M. G., 2017. Species richness alone does not predict cultural ecosystem service value. Proceedings of the National Academy of Sciences of the United States of America, 114(14), 3774–3779. https://doi.org/10.1073/pnas.1701370114

Green, J., 2018. A Dashboard For The Earth. Huffpost. Retrieved from https://www.huffpost.com/entry/a-dashboard-for-the-earth_b_14256028?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAFvhxJBMtIjiJj3WcrohSDWWiw8StwfK-j_a9INZQhRmJUbk3M47ViW5oX3xnX7sHC_mqGR6WSfAYuI0HCSJO3ftjx0hF1PLA4-f-wXa1m

Greenspace Information for Greater London CIC, 2009. Habitat Action Plans. Accessed on 7 February 2024 at: https://www.gigl.org.uk/about-gigl/planning-projects/londons-biodiversity-action-plan/london-bap-priority-habitats/

Greenspace Information for Greater London CIC, 2023a. GLA 10 Minute Walk map. Accessed on 7 February 2024 at: https://www.gigl.org.uk/2023/02/16/gla-10-minute-walk-map/

Greenspace Information for Greater London CIC, 2023b. Key facts and figures from GiGL data holdings. Accessed on 7 February 2024 at: https://www.gigl.org.uk/our-data-holdings/keyfigures/

Greenspace Information for Greater London CIC, 2023c. Open Space Data. Description of the dataset can be found at: https://www.gigl.org.uk/our-data-holdings/open-spaces/

Greenspace Information for Greater London CIC, 2023d. Species Records for London's Priority Species List. Description of dataset can be found at: https://www.gigl.org.uk/about-gigl/planning-projects/londons-biodiversity-action-plan/london-bap-priority-species/

Griffin, A., Stewart, L., and Young, A., 2019. Revising the BFIHOST catchment descriptor to improve UK flood frequency estimates. Hydrology Research, 50(6), 1508–1519. https://doi.org/10.2166/nh.2019.166

Guedes-Santos, J., Correia, R. A., Jepson, P., and Ladle, R., 2021. Evaluating public interest in protected areas using Wikipedia page views. Journal for Nature Conservation, 63, 126040.

Handley, J., Pauleit, S., Slinn, P., Lindley, S., A, M. B., Barber, L., and Jones, C., 2003. Providing accessible natural greenspace in towns and cities: a practical guide to assessing the resource and implementing local standards for provision. Retrieved from http://publications.naturalengland.org.uk/file/78003

Häfner, K., Zasada, I., van Zanten, B. T., Ungaro, F., Koetse, M., and Piorr, A., 2018. Assessing landscape preferences: a visual choice experiment in the agricultural region of Märkische Schweiz, Germany. Landscape Research, 43(6), 846–861. https://doi.org/10.1080/01426397.2017.1386289

Hamel, P., Guerry, A. D., Polasky, S., Han, B., Douglass, J. A., Hamann, M., … Daily, G. C., 2021. Mapping the benefits of nature in cities with the InVEST software. Urban Sustainability, 1(1). https://doi.org/10.1038/s42949-021-00027-9

Hampton, S. E., Strasser, C. A., Tewksbury, J. J., Gram, W. K., Budden, A. E., Batcheller, A. L. et al., 2013. Big data and the future of ecology. Frontiers in Ecology and the Environment, 11(3), 156–162. https://doi.org/10.1890/120103

Hahs, A. K., McDonnell, M. J., McCarthy, M. A., Vesk, P. A., Corlett, R. T., Norton, B. A. et al., 2009. A global synthesis of plant extinction rates in urban areas. Ecology Letters, 12(11), 1165–1173. https://doi.org/10.1111/j.1461-0248.2009.01372.x

Hanski, I., 2015. Habitat fragmentation and species richness. Journal of Biogeography, 42(5), 989–993. https://doi.org/10.1111/jbi.12478

Hanson, H. I., Eckberg, E., Widenberg, M., and Alkan Olsson, J., 2021. Gardens' contribution to people and urban green space. Urban Forestry and Urban Greening, 63(May), 127198. https://doi.org/10.1016/j.ufug.2021.127198

Harrison, P., Sier, A., Acreman, M., Bealey, B., Fry, M., Jones, L., et al., 2017. Natural Capital Metrics. Phase 1 Final Report: Central components. CEH Project NEC06063. Retrieved from: https://www.ceh.ac.uk/sites/default/files/documents/Natural_Capital_Metrics_project_NEC06063_Final_Report_central_components.pdf

Hassani, H., and Silva, E. S., 2023. The Role of ChatGPT in Data Science : How AI-Assisted Conversational Interfaces Are Revolutionizing the Field.

Hastie, T., Tibshirani, R., and Friedman, J., 2017. The Elements of Statistical Learning The Elements of Statistical Learning. Retrieved from https://web.stanford.edu/~hastie/Papers/ESLII.pdf

Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Kulkarni, R., Tenkanen, H., and Di Minin, E., 2020. Understanding sentiment of national park visitors from social media data. People and Nature, 2(3), 750–760. https://doi.org/10.1002/pan3.10130

Havinga, I., Marcos, D., Bogaart, P., Massimino, D., Hein, L., and Tuia, D., 2023. Social media and deep learning reveal specific cultural preferences for biodiversity. People and Nature, 5(3), 981–998. https://doi.org/10.1002/pan3.10466

Heaviside, C., Macintyre, H., and Vardoulakis, S., 2017. The Urban Heat Island: Implications for Health in a Changing Environment. Current Environmental Health Reports, 4(3), 296–305. https://doi.org/10.1007/s40572-017-0150-3

Hernández-Morcillo, M., Plieninger, T., and Bieling, C., 2013. An empirical review of cultural ecosystem service indicators. Ecological Indicators, 29, 434–444. https://doi.org/10.1016/j.ecolind.2013.01.013

Hijmans, R., 2022. Terra: Spatial Data Analysis. R package version 1.6-17, https://CRAN.R-project.org/package=terra

Hinson, C., O'Keeffe, J., Mijic, A., Bryden, J., Van Grootveld, J., and Collins, A. M., 2022. Using natural capital and ecosystem services to facilitate participatory environmental decision making: Results from a systematic map. People and Nature, 4(3), 652–668. https://doi.org/10.1002/pan3.10317

Holt, A.R., 2017. Natural capital assessments for Braithwaite, Glenridding and Staveley sub-catchments. Cumbria Catchment Pioneer Pilot Project Phase 1. Natural Capital Solutions. Retrieved from: https://ecosystemsknowledge.net/sites/default/files/wp-content/uploads/Cumbria-Catchment-Pioneer-Pilot_phases-1_v2.compressed.pdf

Hornigold, K., Lake, I., and Dolman, P., 2016. Recreational use of the countryside: No evidence that high nature value enhances a key ecosystem service. PLoS ONE, 11(11), 1–14. https://doi.org/10.1371/journal.pone.0165043

Howden, N. J. K., Burt, T. P., Worrall, F., Whelan, M. J., and Bieroza, M., 2010. Nitrate concentrations and fluxes in the River Thames over 140 years (1868-2008): Are increases irreversible? Hydrological Processes, 24(18), 2657–2662. https://doi.org/10.1002/hyp.7835

Hungate, B. A., and Cardinale, B. J., 2017. Biodiversity: what value should we use? Frontiers in Ecology and the Environment, 15(6), 283. https://doi.org/10.1002/fee.1511

IPBES, 2019. Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services - Advance Unedited Version – assessment Key messages. Retrieved from https://www.dropbox.com/sh/yd8l2v0u4jqptp3/AACpraYjOYWpTxAFv5H-2vrKa/1 Global Assessment Summary for Policymakers?dl=0&subfolder_nav_tracking=1

IPBES, 2022. Methodological Assessment Report on the Diverse Values and Valuation of Nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem

Services. (D. P. Balvanera, P. Balvanera, U. Pascual, C. Michael, B. Baptiste, González-Jiménez, ed.). Retrieved from https://doi.org/10.5281/zenodo.6418971

Isaac, N. J. B., Brotherton, P. N. M., Bullock, J. M., Gregory, R. D., Boehning-Gaese, K., Connor, B. et al., 2018. Defining and delivering resilient ecological networks: Nature conservation in England. Journal of Applied Ecology. https://doi.org/10.1111/1365-2664.13196

Isaac, N. J. B., and Pocock, M. J. O., 2015. Bias and information in biological records. Biological Journal of the Linnean Society, 115(3), 522–531. https://doi.org/10.1111/bij.12532

Isaac, N. J. B., van Strien, A. J., August, T. A., de Zeeuw, M. P., and Roy, D. B., 2014. Statistics for citizen science: Extracting signals of change from noisy ecological data. Methods in Ecology and Evolution, 5(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254

Jaganmohan, M., Knapp, S., Buchmann, C. M., and Schwarz, N., 2016. The Bigger, the Better? The Influence of Urban Green Space Design on Cooling Effects for Residential Areas. Journal of Environmental Quality, 45(1), 134–145. https://doi.org/10.2134/jeq2015.01.0062

Jaligot, R., Hasler, S., and Chenal, J., 2019. National assessment of cultural ecosystem services: Participatory mapping in Switzerland. Ambio, 48(10), 1219–1233.

Jarvie, H. P., Neal, C., and Withers, P. J. A., 2006. Sewage-effluent phosphorus: A greater risk to river eutrophication than agricultural phosphorus ? 360, 246–253. https://doi.org/10.1016/j.scitotenv.2005.08.038

Jarvie, H. P., Smith, D. R., Norton, L. R., Edwards, F. K., Bowes, M. J., King, S. M. et al., 2018. Phosphorus and nitrogen limitation and impairment of headwater streams relative to rivers in Great Britain: A national perspective on eutrophication. Science of the Total Environment, 621(December), 849–862. https://doi.org/10.1016/j.scitotenv.2017.11.128

Jarvis, A., Reuter, H.I., Nelson, A., and Guevara, E., 2008. Hole-filled seamless SRTM data V4. Retrieved from http://srtm.csi.cgiar.org

Jasiewicz, J., 2021. r.stream.snap - Snap point to modelled stream network. https://grass.osgeo.org/grass82/manuals/addons/r.stream.snap.html

Johnes, P. J., 1996. Evaluation and management of the impact of land use change on the nitrogen and phosphorus load delivered to surface waters: The export coefficient modelling approach. Journal of Hydrology, 183(3–4), 323–349. https://doi.org/10.1016/0022-1694(95)02951-6

Johnes, P. J., Heathwaite, A. L., Spears, B. M., Brownlie, W. J., Elser, J. J., Haygarth, P. M. et al., 2022. Phosphorus and water quality. Our Phosphorus Future, (May). https://doi.org/10.13140/RG.2.2.14950.50246

Jones, L., Norton, L., Austin, Z., Browne, A. L., Donovan, D., Emmett, B. A. et al., 2016. Stocks and flows of natural and human-derived capital in ecosystem services. Land Use Policy, 52, 151–162. https://doi.org/10.1016/j.landusepol.2015.12.014

Jones, L., Vieno, M., Fitch, A., Carnell, E., Steadman, C., Cryle, P. et al., 2019. Urban natural capital accounts: developing a novel approach to quantify air pollution removal by vegetation. Journal of Environmental Economics and Policy, 8(4), 413–428. https://doi.org/10.1080/21606544.2019.1597772

Jordan, M. I., and Mitchell, T. M., 2015. Machine learning: Trends, perspectives, and prospects. Science, 349(6245).

Kareiva P., Tallis H., Ricketts T.H., Daily G.C. and Polasky S. (eds.), 2011. Natural Capital: theory and practice of mapping ecosystem services. Oxford University Press, Oxford.

Karimi, J. D., Corstanje, R., and Harris, J. A., 2021. Understanding the importance of landscape configuration on ecosystem service bundles at a high resolution in urban landscapes in the UK. Landscape Ecology, 36(7), 2007–2024. https://doi.org/10.1007/s10980-021-01200-2

Kassambara, A., and Mundt, F., 2020. factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7, <https://CRAN.R-project.org/package=factoextra>.

Kassambara, A., 2023. ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'. R package version 0.1.4.1, <https://CRAN.R-project.org/package=ggcorrplot>.

Kattenborn T., Leitloff J., Schiefer F. and Hinz S., 2021. Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173: 24-49.

Katz-Gerro, T., and Orenstein, D. E., 2015. Environmental tastes, opinions and behaviors: Social sciences in the service of cultural ecosystem service assessment. Ecology and Society, 20(3). https://doi.org/10.5751/ES-07545-200328

Kelemen, E., García-Llorente, M., Pataki, G., Martín-López, B., and Gómez-Baggethun, E., 2016. Non-monetary techniques for the valuation of ecosystem services. In M. Potschin and K. Jax (Eds.), OpenNESS Ecosystem Services Reference Book. Retrieved from www.openness- project.eu/library/reference-book%0AAcknowledgements:

Kéry, M., Dorazio, R. M., Soldaat, L., Van Strien, A., Zuiderwijk, A., and Royle, J. A., 2009. Trend estimation in populations with imperfect detection. Journal of Applied Ecology, 46(6), 1163–1172. https://doi.org/10.1111/j.1365-2664.2009.01724.x

Khadivi, P., and Ramakrishnan, N., 2016. Wikipedia in the Tourism Industry : Forecasting Demand and Modeling Usage Behavior. Proceedings of the AAAI Conference on Artificial Intelligence, 30(2), 4016–4021. Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/19078

Kong, F., Yin, H., Wang, C., Cavan, G., and James, P., 2014. A satellite image-based analysis of factors contributing to the green-space cool island intensity on a city scale. Urban Forestry and Urban Greening, 13(4), 846–853. https://doi.org/10.1016/j.ufug.2014.09.009

Ladle, R. J., Correia, R. A., Do, Y., Joo, G. J., Malhado, A. C. M., Proulx, R. et al., 2016. Conservation culturomics. Frontiers in Ecology and the Environment, 14(5), 269–275. https://doi.org/10.1002/fee.1260

Ladle, R. J., Jepson, P., Correia, R. A., and Malhado, A. C. M., 2019. A culturomics approach to quantifying the salience of species on the global internet. People and Nature, 1(4), 524–532. https://doi.org/10.1002/pan3.10053

Landau, V.A., Shah, V.B., Anantharaman, R., and Hall, K.R., 2021. Omniscape.jl: Software to compute omnidirectional landscape connectivity. Journal of Open Source Software, 6(57), 2829.

Langhans, K. E., Schmitt, R. J. P., Chaplin-Kramer, R., Anderson, C. B., Vargas Bolaños, C., Vargas Cabezas, F. et al. 2022. Modeling multiple ecosystem services and beneficiaries of riparian reforestation in Costa Rica. Ecosystem Services, 57(October 2021). https://doi.org/10.1016/j.ecoser.2022.101470

Lavallin, A., and Downs, J. A., 2021. Machine learning in geography–Past, present, and future. Geography Compass, 15(5). https://doi.org/10.1111/gec3.12563

Lawton, J. H., Brotherton, P. N. M. P. N. M., Brown, V. K. K., Elphick, C., Fitter, A. H., Forshaw, J. et al., 2010. Making space for nature: a review of England's wildlife sites and ecological network. Report to Defra, (September), 1–107. Retrieved from https://webarchive.nationalarchives.gov.uk/20130402170324/http:/archive.defra.gov.uk/environment/biodiversity/documents/201009space-for-nature.pdf

Lear, R., Wigley, S., Lord, A., Lusardi, J., and Rice, P., 2021. Natural Capital Atlases: Mapping Indicators for County and City Regions - NECR318. Retrieved from http://publications.naturalengland.org.uk/publication/6672365834731520

Li, L., Goodchild, M. F., and Xu, B., 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartography and Geographic Information Science, 40(2), 61–77.

Li, Y., Zhang, H., Xue, X., Jiang, Y., and Shen, Q., 2018. Deep learning for remote sensing image classification: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(6), 1–17. https://doi.org/10.1002/widm.1264

Li, X., and Zhou, W., 2019. Optimizing urban greenspace spatial pattern to mitigate urban heat island effects: Extending understanding from local to the city scale. Urban Forestry and Urban Greening, 41(October 2018), 255–263. https://doi.org/10.1016/j.ufug.2019.04.008

Liaw, A and Wiener, M., 2002. Classification and Regression by randomForest. R News 2(3), 18--22. https://CRAN.R-project.org/package=randomForest

Lin, Y.-P., Lin, W.-C., Lien, W.-Y., Anthony, J., and Petway, J., 2017. Identifying Reliable Opportunistic Data for Species Distribution Modeling: A Benchmark Data Optimization Approach. Environments, 4(4), 81. https://doi.org/10.3390/environments4040081

Liu, H., Remme, R. P., Hamel, P., Nong, H., & Ren, H., 2020. Supply and demand assessment of urban recreation service and its implication for greenspace planning-A case study on Guangzhou. Landscape and Urban Planning, 203(October 2019), 103898. https://doi.org/10.1016/j.landurbplan.2020.103898

Liu, P., Di, L., Du, Q., and Wang, L., 2018. Remote sensing big data: Theory, methods and applications. Remote Sensing, 10(5), 2016–2019. https://doi.org/10.3390/rs10050711

Lockwood, J. L., Cassey, P., and Blackburn, T. M., 2009. The more you introduce the more you get: The role of colonization pressure and propagule pressure in invasion ecology. Diversity and Distributions, 15(5), 904–910. https://doi.org/10.1111/j.1472-4642.2009.00594.x

London Councils, 2018. A London Councils Member briefing Valuing London's Parks and Green Spaces Overview. Retrieved from www.londoncouncils.gov.uk/members-area

Lucas, T. C. D., 2020. A translucent box: interpretable machine learning in ecology. Ecological Monographs, 90(4), 1–17. https://doi.org/10.1002/ecm.1422

Luna Juncal, M. J., Masino, P., Bertone, E., and Stewart, R. A., 2023. Towards nutrient neutrality: A review of agricultural runoff mitigation strategies and the development of a decision-making framework. Science of the Total Environment, 874(February), 162408. https://doi.org/10.1016/j.scitotenv.2023.162408

Mace, G. M., 2014. Whose conservation? Science, 345(6204), 1558–1560. https://doi.org/10.1126/science.1254704

Mace, G. M., 2019. The ecology of natural capital accounting. Oxford Review of Economic Policy, 35(1), 54–67. https://doi.org/10.1093/oxrep/gry023

Mace, G. M., Hails, R. S., Cryle, P., Harlow, J., and Clarke, S. J., 2015. Towards a risk register for natural capital. Journal of Applied Ecology, 52(3), 641–653. https://doi.org/10.1111/1365-2664.12431

Maddock, A., 2011. UK Biodiversity Action Plan: Priority Habitat Descriptions. JNCC, Peterborough. Retrieved from: https://data.jncc.gov.uk/data/2728792c-c8c6-4b8c-9ccd-a908cb0f1432/UKBAP-PriorityHabitatDescriptions-Rev-2011.pdf

Maes, D., Isaac, N. J. B., Harrower, C. A., Collen, B., van Strien, A. J., and Roy, D. B., 2015. The use of opportunistic data for IUCN Red List assessments. Biological Journal of the Linnean Society, 115(3), 690–706. https://doi.org/10.1111/bij.12530

Maes, J., Egoh, B., Willemen, L., Liquete, C., Vihervaara, P., Schägner, J. P. et al., 2012. Mapping ecosystem services for policy support and decision making in the European Union. Ecosystem Services, 1(1), 31–39. https://doi.org/10.1016/j.ecoser.2012.06.004

Maestrini, B., Mimić, G., van Oort, P. A. J., Jindo, K., Brdar, S., van Evert, F. K., and Athanasiados, I., 2022. Mixing process-based and data-driven approaches in yield prediction. European Journal of Agronomy, 139(August 2021). https://doi.org/10.1016/j.eja.2022.126569

Mahendra, H. N., Mallikarjunaswamy, S., Siddesh, G. K., Komala, M., and Sharmila, N. 2020. Evolution of real-time onboard processing and classification of remotely sensed data.

Indian Journal of Science and Technology, 13(20), 2010–2020. https://doi.org/10.17485/ijst/v13i20.459

Mainstone, C. P., and Parr, W., 2002. Phosphorus in rivers - Ecology and management. Science of the Total Environment, 282–283, 25–47. https://doi.org/10.1016/S0048-9697(01)00937-8

Mancini, F., Coghill, G. M., and Lusseau, D., 2018. Using social media to quantify spatial and temporal dynamics of nature-based recreational activities. PLoS ONE, 13(7), 1–19. https://doi.org/10.1371/journal.pone.0200565

Mancini, F., Coghill, G. M., and Lusseau, D., 2019. Quantifying wildlife watchers' preferences to investigate the overlap between recreational and conservation value of natural areas. Journal of Applied Ecology, 56(2), 387–397. https://doi.org/10.1111/1365-2664.13274

Mancini, F., Hodgson, J. A., and Isaac, N. J. B., 2022. Co-designing an Indicator of Habitat Connectivity for England. Frontiers in Ecology and Evolution, 10(July), 1–12. https://doi.org/10.3389/fevo.2022.892987

Marston, C. G., O'Neil, A. W., Morton, R. D., Wood, C. M., and Rowland, C. S., 2023. LCM2021 -- the UK Land Cover Map 2021. Earth System Science Data, 15(10), 4631–4649. https://doi.org/10.5194/essd-15-4631-2023

Mata, L., Ramalho, C. E., Kennedy, J., Parris, K. M., Valentine, L., Miller, M. et al., 2020. Bringing nature back into cities. People and Nature, 2(2), 350–368. https://doi.org/10.1002/pan3.10088

Mavromatis, T., 2014. Pre-season prediction of regional rainfed wheat yield in Northern Greece with CERES-Wheat. Theor. Appl. Climatol. 117, 653–665. https://doi.org/ 10.1007/s00704-013-1031-9.

Maxwell, A. E., Warner, T. A., and Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. International Journal of Remote Sensing, 39(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Mayor of London, 2017. Natural Capital Accounts for Public Green Space in London. 35. Retrieved from https://www.london.gov.uk/sites/default/files/11015viv_natural_capital_account_for_london_v7_full_vis.pdf

McCauley, D. J., 2006. Selling out on nature. Nature, 443(7107), 27–28. https://doi.org/10.1038/443027a

McCrea, R., King, R., Graham, L., and Börger, L., 2023. Realising the promise of large data and complex models. Methods in Ecology and Evolution, 14(1), 4–11. https://doi.org/10.1111/2041-210X.14050

McGarigal, K., Cushman, S.A. and Ene, E., 2012. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps. Computer Software Program Produced by the Authors at the University of Massachusetts, Amherst. http://www.umass.edu/landeco/research/fragstats/fragstats.html

McGarigal K., Cushman, S.A., and Ene, E., 2023. FRAGSTATS v4: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors; available at the following web site: https://www.fragstats.org

McRae, B. H., Popper, K., Jones, A., Schindel, M., Buttrick, S., Hall, K. R., Unnasch, R. S., and Platt, J., 2016. Conserving Nature's Stage: Mapping Omnidirectional Connectivity for Resilient Terrestrial Landscapes in the Pacific Northwest. The Nature Conservancy, Portland, Oregon.

Meneses-Tovar, C. L., 2011. NDVI as indicator of degradation. Unasylva, 62(238), 39–46.

Met Office, 2020: HadUK-Grid Gridded Climate Observations on a 1km grid over the UK, v1.0.2.1 (1862-2019). Centre for Environmental Data Analysis, 21 October 2020. doi:10.5285/89908dfcb97b4a28976df806b4818639. http://dx.doi.org/10.5285/89908dfcb97b4a28976df806b4818639

Mets, K. D., Armenteras, D., and Dávalos, L. M., 2017. Spatial autocorrelation reduces model precision and predictive power in deforestation analyses. Ecosphere, 8(5). https://doi.org/10.1002/ecs2.1824

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P. et al., 2011. Quantitative Analysis of Culture Using Millions of Digitized Books. Science, 331(6014), 176–182. https://doi.org/10.1126/science.1199644

Milcu, A. I., Hanspach, J., Abson, D., and Fischer, J., 2013. Cultural ecosystem services: A literature review and prospects for future research. Ecology and Society, 18(3). https://doi.org/10.5751/ES-05790-180344

Millard, J. W., Gregory, R. D., Jones, K., and Freeman, R., 2021. The species awareness index as a conservation culturomics metric for public biodiversity awareness. Conservation Biology, 35(2), 472–482. https://doi.org/10.1111/cobi.13701

Millenium Ecosystem Assessment, 2005. Living Beyond Our Means: Natural Assets and Human Well-Being. Retrieved from https://www.millenniumassessment.org/documents/document.429.aspx.pdf

Missemer, A., 2018. Natural Capital as an Economic Concept, History and Contemporary Issues. Ecological Economics, 143, 90–96. https://doi.org/10.1016/j.ecolecon.2017.07.011

Mitchell, M. G. E., Suarez-Castro, A. F., Martinez-Harms, M., Maron, M., McAlpine, C., Gaston, K. J. et al., 2015. Reframing landscape fragmentation's effects on ecosystem services. Trends in Ecology and Evolution, 30(4), 190–198. https://doi.org/10.1016/j.tree.2015.01.011

Mittermeier, J. C., Correia, R., Grenyer, R., Toivonen, T., and Roll, U., 2021a. Using Wikipedia to measure public interest in biodiversity and conservation. Conservation Biology, 35(2), 412–423. https://doi.org/10.1111/cobi.13702

Mittermeier, J. C., Roll, U., Matthews, T. J., Correia, R., and Grenyer, R., 2021b. Birds that are more commonly encountered in the wild attract higher public interest online. Conservation Science and Practice, 3(5), 1–11. https://doi.org/10.1111/csp2.340

Molnar, C., 2023. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd Edition, https://christophm.github.io/interpretable-ml-book/

Moore, R.V., Morris, D.G. and Flavin, R.W., 1994. Sub-set of UK digital 1:50,000 scale river centre-line network. NERC, Institute of Hydrology, Wallingford.

Moorhouse, H. L., McGowan, S., Taranu, Z. E., Gregory-Eaves, I., Leavitt, P. R., Jones, M. D. et al., 2018. Regional versus local drivers of water quality in the Windermere catchment, Lake District, United Kingdom: The dominant influence of wastewater pollution over the past 200 years. Global Change Biology, 24(9), 4009–4022. https://doi.org/10.1111/gcb.14299

Morris, D.G., and Flavin, R.W., 1990. A digital terrain model for hydrology. Proc 4th International Symposium on Spatial Data Handling. Vol 1 Jul 23-27 Zurich, pp 250-262.

Morris, D. G., and Flavin, R. W., 1994. Sub-set of UK 50 m by 50 m hydrological digital terrain model grids. NERC, Institute of Hydrology, Wallingford.

Morton, R.D., Marston, C.G., O'Neil, A.W., Rowland, C.S., 2020. Land Cover Map 2017 (25m rasterised land parcels, GB). NERC Environmental Information Data Centre. https://doi.org/10.5285/499212cd-d64a-43ba-b801-95402e4d4098

Morton, R. D., and Rowland, C. S., 2015. Developing and Evaluating an Earth Observation-enabled ecological land cover time series system. JNCC Report No : 563, (June).

Moss, B., 2008. Water pollution by agriculture. Philos Trans R Soc Lond B Biol Sci. Feb 12;363:659-66. doi: 10.1098/rstb.2007.2176.

Muñoz, L., Hausner, V. H., Runge, C., Brown, G., and Daigle, R., 2020. Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. People and Nature, 2(2), 437–449. https://doi.org/10.1002/pan3.10083

Mustika, P. L. K., Ichsan, M., and Booth, H., 2020. The Economic Value of Shark and Ray Tourism in Indonesia and Its Role in Delivering Conservation Outcomes. Frontiers in Marine Science, 7(April). https://doi.org/10.3389/fmars.2020.00261

Nassauer, J. I., 1983. Framing the Landscape in Photographic Simulation. Journal of Environmental Management, 17, 1-16.

Nassauer, J. I., Wang, Z., and Dayrell, E., 2009. What will the neighbors think? Cultural norms and ecological design. Landscape and Urban Planning, 92(3–4), 282–292.

Natural Capital Coalition, 2016. The path towards the Natural Capital Protocol: a primer for business. Retrieved from https://naturalcapitalcoalition.org/wp-content/uploads/2016/07/NCC_Primer_WEB_2016-07-08.pdf

Natural Capital Committee, 2013. The State of Natural Capital: Towards a framework for measurement and valuation. (April).

Natural Capital Committee, 2014a. Towards a Framework for Defining and Measuring Changes in Natural Capital. (March), 1–21.

Natural Capital Committee, 2014b. The State of Natural Capital: Restoring our Natural Assets. Retrieved from www.naturalcapitalcommittee.org.

Natural Capital Committee, 2019. State of Natural Capital Annual Report 2019. 69. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/774218/ncc-annual-report-2019.pdf

Natural Capital Committee, 2020. End of Term Report. To the domestic and economy implementation committee of the cabinet. 56. Available at: https://assets.publishing.service.gov.uk/media/5f9c3befe90e070428701525/ncc-end-of-term-report.pdf

Natural England, 2018. Full report Natural Capital Indicators: for defining and measuring change in natural capital Project manager. Retrieved from: https://publications.naturalengland.org.uk/publication/6742480364240896

Natural England, 2021. Greater London Natural Capital Atlas: Mapping Indicators. https://publications.naturalengland.org.uk/publication/6672365834731520

Natural England, 2024. Natural England's Priority Habitat Inventory V2.3. Retrieved from: https://naturalengland-defra.opendata.arcgis.com/datasets/Defra::priority-habitats-inventory-england/about

Natural Environment and Rural Communities Act, 2006, Part 3, Biodiversity, Section 41. Retrieved from: https://www.legislation.gov.uk/ukpga/2006/16/section/41

Neal, C., Jarvie, H. P., Withers, P. J. A., Whitton, B. A., and Neal, M., 2010. The strategic significance of wastewater sources to pollutant phosphorus levels in English rivers and to environmental management for rural, agricultural and urban catchments. Science of the Total Environment, 408(7), 1485–1500. https://doi.org/10.1016/j.scitotenv.2009.12.020

Neel, M. C., McGarigal, K., and Cushman, S. A., 2004. Behavior of class-level landscape metrics across gradients of class aggregation and area. Landscape Ecology, 19(May 2004), 435–455. https://doi.org/10.1023/B

Neumann, J. L., Griffiths, G. H., Hoodless, A., & Holloway, G. J., 2016. The compositional and configurational heterogeneity of matrix habitats shape woodland carabid communities in wooded-agricultural landscapes. Landscape Ecology, 31(2), 301–315. https://doi.org/10.1007/s10980-015-0244-y

Nisbet, T. R., Andreucci, M.-B., De Vreese, R., Högbom, L., Kay, S., Kelly-Quinn, M. et al., 2022. Forest Green Infrastructure to Protect Water Quality: A Step-by-Step Guide for Payment Schemes. 105–131. https://doi.org/10.1007/978-981-16-6791-6_8

Nolan, G., Kane, A., and Fernández-Bellon, D., 2022. Natural history films generate more online interest in depicted species than in conservation messages. People and Nature, 4(3), 816–825. https://doi.org/10.1002/pan3.10319

Norgaard, R. B., 2010. Ecosystem services: From eye-opening metaphor to complexity blinder. Ecological Economics, 69(6), 1219–1227. https://doi.org/10.1016/j.ecolecon.2009.11.009

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., and Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences of the United States of America, 115(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Norton, B. A., Evans, K. L., & Warren, P. H., 2016. Urban Biodiversity and Landscape Ecology: Patterns, Processes and Planning. Current Landscape Ecology Reports, 1(4), 178–192. https://doi.org/10.1007/s40823-016-0018-5

OECD, 2023. Inflation (CPI). Available at: https://data.oecd.org/price/inflation-cpi.htm Accessed: 26/03/2024

Office for National Statistics, 2011. Population density [GeoPackage geospatial data], Scale 1:2500, Tiles: GB, Updated: 27 March 2011, ONS, Using: EDINA Society Digimap Service, <https://digimap.edina.ac.uk>, Downloaded: 2021-06-24 16:03:58.872

Office for National Statistics, 2015. Major Towns and Cities (December 2015) Boundaries V2. Retrieved from https://geoportal.statistics.gov.uk/datasets/ons::major-towns-and-cities-december-2015-boundaries-v2/about

Office for National Statistics, 2019. Countries (December 2019) Boundaries UK BFE. Retrieved from https://geoportal.statistics.gov.uk/datasets/countries-december-2019-boundaries-uk-bfe?geometry=-31.281%2C51.101%2C26.419%2C59.782

Office for National Statistics, 2020. One in eight British households has no garden, accessed 07 February 2024 at: https://www.ons.gov.uk/economy/environmentalaccounts/articles/oneineightbritishhouseholdshasnogarden/2020-05-14

Office for National Statistics, 2022a. Woodland natural capital accounts 2022. Retrieved from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/woodlandnaturalcapitalaccountsuk/2022

Office for National Statistics, 2022b. Health benefits from recreation, natural capital, UK: 2022. 2–22. Retrieved from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/healthbenefitsfromrecreationnaturalcapitaluk/2022

Office for National Statistics, 2022c. UK natural capital accounts: 2022. Office for National Statistics, (March), 1–18. Retrieved from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/uknaturalcapitalaccounts/2022

Office for National Statistics, 2023a. UK natural capital accounts: 2023, https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/uknaturalcapitalaccounts/2023

Office for National Statistics, 2023b. Urban natural capital accounts, UK. (December), 1–27. Retrieved from https://www.ons.gov.uk/economy/environmentalaccounts/bulletins/urbannaturalcapitalaccountsuk/2023

O'Hare, M. T., Baattrup-Pedersen, A., Baumgarte, I., Freeman, A., Gunn, I. D. M., Lázár, A. N., Sinclair, R., Wade, A. J., Bowes, M. J., 2018. Responses of aquatic plants to eutrophication in rivers: A revised conceptual model. Frontiers in Plant Science, 9(April), 1–13. https://doi.org/10.3389/fpls.2018.00451

O'Keeffe, J., Pluchinotta, I., De Stercke, S., Hinson, C., Puchol-Salort, P., Mijic, A. et al., 2022. Evaluating natural capital performance of urban development through system dynamics: A case study from London. Science of the Total Environment, 824, 153673. https://doi.org/10.1016/j.scitotenv.2022.153673

Oliver, R. Y., Ellis, D. P. W., Chmura, H. E., Krause, J. S., Pérez, J. H., Sweet, S. K. et al., 2018. Eavesdropping on the Arctic: Automated bioacoustics reveal dynamics in songbird breeding phenology. Science Advances, 4(6), 1–9. https://doi.org/10.1126/sciadv.aaq1084

OpenStreetMap, 2020. OpenStreetMap Data in Layered GIS Format. Retrieved from http://download.geofabrik.de/europe/great-britain/england-latest-free.shp.zip

OpenStreetMap, 2024. OpenStreetMap Data in Layered GIS Format. Retrieved from http://download.geofabrik.de/europe/great-britain/england-latest-free.shp.zip

Ordnance Survey Limited, 2022. OS MasterMap Greenspace Layer, accessed via Digimap, https://www.ordnancesurvey.co.uk/products/os-mastermap-greenspace-layer

Pan, Y., and Vira, B., 2019. Exploring natural capital using bibliometrics and social media data. Ecology and Society, 24(4). https://doi.org/10.5751/ES-11118-240405

Panda, S. S., Ames, D. P., and Panigrahi, S., 2010. Application of vegetation indices for agricultural crop yield prediction using neural network techniques. Remote Sensing, 2(3), 673–696. https://doi.org/10.3390/rs2030673

Park, C. Y., Choe, H., and Thorne, J. H., 2023. Mapping urban cool air connectivity in a megacity. Urban Climate, 47(June 2022), 101362. https://doi.org/10.1016/j.uclim.2022.101362

Parliamentary Office of Science and Technology, 2011. POSTnote 376: Natural Capital Accounting. Retrieved from https://post.parliament.uk/research-briefings/post-pn-376/

Pearson, R. G., 2016. Reasons to Conserve Nature. Trends in Ecology and Evolution, 31(5), 366–371. https://doi.org/10.1016/j.tree.2016.02.005

Perry, G. L. W., Seidl, R., Bellvé, A. M., and Rammer, W., 2022. An Outlook for Deep Learning in Ecosystem Science. Ecosystems, 25(8), 1700–1718. https://doi.org/10.1007/s10021-022-00789-y

Peterson, G. D., Harmáčková, Z. V., Meacham, M., Queiroz, C., Jiménez-Aceituno, A., Kuiper, J. J. et al., 2018. Welcoming different perspectives in IPBES: “nature’s contributions to people” and “ecosystem services.” Ecology and Society, 23(1), 39. https://doi.org/10.5751/ES-10134-230139

Pettorelli, N., Owen, H. J. F., and Duncan, C., 2016. How do we want Satellite Remote Sensing to support biodiversity conservation globally? Methods in Ecology and Evolution, 7(6), 656–665. https://doi.org/10.1111/2041-210X.12545

Pettorelli, N., Williams, J., Schulte to Bühne, H., Crowson, M., in press. Deep learning and satellite remote sensing for biodiversity monitoring and conservation. Remote Sensing in Ecology and Conservation.

Phinney, S., 2022. Accounting for Natural Capital in Cities: Making the Invisible Visible. https://www.liverpool.ac.uk/heseltine-institute/publications/

Pichler, M., and Hartig, F., 2023. Machine learning and deep learning—A review for ecologists. Methods in Ecology and Evolution, 14(4), 994–1016. https://doi.org/10.1111/2041-210X.14061

Pinard, M. A., Putz, F. E., Tay, J., and Sullivan, T. E., 1995. Guidelines for a Reduced-Impact Logging Project in Malaysia. Journal of Forestry, 41–45.

Planet Labs PBC, 2022. Planet Application Program Interface: In Space for Life on Earth, "https://api.planet.com"

Popkin, G., 2016. Satellite alerts track deforestation in real time. Nature, 530(7591), 392–393. https://doi.org/10.1038/530392a

Posner, S. M., McKenzie, E., and Ricketts, T. H., 2016. Policy impacts of ecosystem services knowledge. Proceedings of the National Academy of Sciences of the United States of America, 113(7), 1760–1765. https://doi.org/10.1073/pnas.1502452113

POSTNote 542, 2016. Natural Capital: An Overview. Retrieved from https://researchbriefings.parliament.uk/ResearchBriefing/Summary/POST-PN-0542

Prasad, R., and Chakraborty, D., 2019. Phosphorus Basics: Understanding Phosphorus Forms and Their Cycling in the Soil. Crop Production, Alabama A&M & Auburn Universities Extension, ANR 2535, 2535. Retrieved from https://www.aces.edu/blog/topics/crop-production/understanding-phosphorus-forms-and-their-cycling-in-the-soil/%0Ahttp://files/852/understanding-phosphorus-forms-and-their-cycling-in-the-soil.html

Pregitzer, C. C., & Bradford, M. A., 2023. Associations between recent land use history and urban forest composition. Urban Forestry and Urban Greening, 90(October), 128134. https://doi.org/10.1016/j.ufug.2023.128134

Proulx, R., Massicotte, P., and Pépino, M., 2014. Googling Trends in Conservation Biology. Conservation Biology, 28(1), 44–51. https://doi.org/10.1111/cobi.12131

Public Health England, 2014. Local action on health inequalities: Improving access to green spaces. (September), 29. https://doi.org/10.1063/1.4972656

Rahman, A. S., and Rahman, A., 2020. Application of principal component analysis and cluster analysis in regional flood frequency analysis: A case study in new South Wales, Australia. Water (Switzerland), 12(3), 1–26. https://doi.org/10.3390/w12030781

R Core Team, 2022. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reaney, S. M., Lane, S. N., Heathwaite, A. L., and Dugdale, L. J., 2011. Risk-based modelling of diffuse land use impacts from rural landscapes upon salmonid fry abundance. Ecological Modelling, 222(4), 1016–1029. https://doi.org/10.1016/j.ecolmodel.2010.08.022

Redhead, J. W., May, L., Oliver, T. H., Hamel, P., Sharp, R., and Bullock, J. M., 2018. National scale evaluation of the InVEST nutrient retention model in the United Kingdom. Science of the Total Environment, 610–611, 666–677. https://doi.org/10.1016/j.scitotenv.2017.08.092

Rees, S. E., Ashley, M., Cameron, A., Mullier, T., Ingle, C., Oates, J. et al., 2022. A marine natural capital asset and risk register—Towards securing the benefits from marine systems and linked ecosystem services. Journal of Applied Ecology, 59(4), 1098–1109. https://doi.org/10.1111/1365-2664.14121

rePLANET, 2024. What Are Biodiversity Credits? Available at: https://www.replanet.org.uk/what-are-biodiversity-credits/ Accessed: 26/03/2024

Richardson, K., Steffen, W., Lucht, W., Bendtsen, J., Cornell, S. E., Donges, J. F., et al., 2023. Earth beyond six of nine planetary boundaries. Science Advances, 9(37), 1–16. https://doi.org/10.1126/sciadv.adh2458

Robin, X., N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. Sanchez, M. Müller. 2011. "pROC: an open-source package for R and S+ to analyze and compare ROC curves." BMC Bioinformatics, 12, 77.

Rouquette, J., 2016. Mapping natural capital and ecosystem services in the Nene Valley. Retrieved from http://www.naturalcapitalsolutions.co.uk/wp-content/uploads/2016/01/Natural-capital-ecosystem-services-in-the-Nene-Valley-summary-report-web-version.pdf

Ross, C. W., Hanan, N. P., Prihodko, L., Anchang, J., Ji, W., and Yu, Q., 2021. Woody-biomass projections and drivers of change in sub-Saharan Africa. Nature Climate Change, 11(5), 449–455. https://doi.org/10.1038/s41558-021-01034-5

Rowcliffe, J. M., and Carbone, C., 2008. Surveys using camera traps: Are we looking to a brighter future? Animal Conservation, 11(3), 185–186. https://doi.org/10.1111/j.1469-1795.2008.00180.x

Rudd, H., Vala, J., and Schaefer, V., 2002. Importance of backyard habitat in a comprehensive biodiversity conservation strategy: A connectivity analysis of urban green spaces. Restoration Ecology, 10(2), 368–375. https://doi.org/10.1046/j.1526-100X.2002.02041.x

Runting, R. K., Phinn, S., Xie, Z., Venter, O., and Watson, J. E. M., 2020. Opportunities for big data in conservation and sustainability. Nature Communications, 11(1), 1–4. https://doi.org/10.1038/s41467-020-15870-0

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., and Ryo, M., 2023. Ten deep learning techniques to address small data problems with remote sensing. International Journal of Applied Earth Observation and Geoinformation, 125(September), 103569. https://doi.org/10.1016/j.jag.2023.103569

Saha, D., Basso, B., Robertson, G.P., 2021. Machine learning improves predictions of agricultural nitrous oxide (N2O) emissions from intensively managed cropping systems. Environ. Res. Lett. 16. https://doi.org/10.1088/1748-9326/abd2f3.

Sarkar, C., Webster, C., Pryor, M., Tang, D., Melbourne, S., Zhang, X., and Jianzheng, L., 2015. Exploring associations between urban green, street design and walking: Results from the Greater London boroughs. Landscape and Urban Planning, 143, 112–125. https://doi.org/10.1016/j.landurbplan.2015.06.013

Schirpke, U., Scolozzi, R., and Tappeiner, U., 2022. Not too small to benefit society: insights into perceived cultural ecosystem services of mountain lakes in the European Alps. Ecology and Society, 27(1). https://doi.org/10.5751/ES-12987-270106

Schreiber, S. G., Schreiber, S., Tanna, R. N., Roberts, D. R., and Arciszewski, T. J., 2022. Statistical tools for water quality assessment and monitoring in river ecosystems – a scoping review and recommendations for data analysis. Water Quality Research Journal, 57(1), 40–57. https://doi.org/10.2166/wqrj.2022.028

Schröter, M., van der Zanden, E. H., van Oudenhoven, A. P. E., Remme, R. P., Serna-Chavez, H. M., de Groot, R. S., and Opdam, P., 2014. Ecosystem Services as a Contested Concept:

A Synthesis of Critique and Counter-Arguments. Conservation Letters, 7(6), 514–523. https://doi.org/10.1111/conl.12091

Schuetz, J. G., and Johnston, A., 2021. Tracking the cultural niches of North American birds through time. People and Nature, 3(1), 251–260. https://doi.org/10.1002/pan3.10173

Scowen, M., Athanasiadis, I. N., Bullock, J. M., Eigenbrod, F., and Willcock, S., 2021. The current and future uses of machine learning in ecosystem service research. Science of the Total Environment, 799, 149263. https://doi.org/10.1016/j.scitotenv.2021.149263

SEDAC, 2018. Gridded Population of the World (GPW), v4. Retrieved from https://sedac.ciesin.columbia.edu/data/collection/gpw-v4/documentation

SEEA, 2013. Implementation Strategy for the System of Environmental-Economic Accounting (SEEA). Retrieved from: https://unstats.un.org/unsd/statcom/44th-session/documents/doc13/BG-SEEA-Implementation-E.pdf

SEEA, 2021. Natural Capital and Ecosystem Services FAQ. Retrieved March from https://seea.un.org/content/natural-capital-and-ecosystem-services-faq#What is natural capital accounting

SEEA, 2023. Frequently Asked Questions. Available at: https://seea.un.org/content/frequently-asked-questions

Sharp, R., Tallis, H. T., Ricketts, T., Guerry, A. D., Wood, S. A., Chaplin-Kramer, R. et al., 2018. InVEST User Guide. Retrieved from http://releases.naturalcapitalproject.org/invest-userguide/latest/

Shen, L. Q., Amatulli, G., Sethi, T., Raymond, P., and Domisch, S., 2020. Estimating nitrogen and phosphorus concentrations in streams and rivers, within a machine learning framework. Scientific Data, 7(1), 1–11. https://doi.org/10.1038/s41597-020-0478-7

Signorelli, S., Reis, F., and Biffignandi, S., 2016. What attracts tourists while planning for a journey? An analysis of three cities utilising Wikipedia page views. 14th Global Forum on Tourism Statistics, (lv), 1–16. Retrieved from http://tsf2016venice.enit.it/images/articles/Papers_Forum/2.4_What attracts tourists while planning for a journey - an analysis of three cities through Wikipedia page views.pdf

Silvertown, J., 2015. Have Ecosystem Services Been Oversold? Trends in Ecology and Evolution, 30(11), 641–648. https://doi.org/10.1016/j.tree.2015.08.007

Smith, A. C., Koper, N., Francis, C. M., and Fahrig, L., 2009. Confronting collinearity: Comparing methods for disentangling the effects of habitat loss and fragmentation. Landscape Ecology, 24(10), 1271–1285. https://doi.org/10.1007/s10980-009-9383-3

Smith, C., Dawson, D., Archer, J., Davies, M., Frith, M., Hughes, E. and Massini, P., 2011. From green to grey; observed changes in garden vegetation structure in London, 1998-2009; London Wildlife Trust, Greenspace Information for Greater London and Greater London Authority. Retrieved from: https://www.rbkc.gov.uk/pdf/garden%20research%20full%20report.pdf

Smith, V. H., and Schindler, D. W., 2009. Eutrophication science: where do we go from here? Trends in Ecology and Evolution, 24(4), 201–207. https://doi.org/10.1016/j.tree.2008.11.009

Soanes, K., Taylor, L., Ramalho, C. E., Maller, C., Parris, K., Bush, J. et al., 2023. Conserving urban biodiversity: Current practice, barriers, and enablers. Conservation Letters, 16(3), 1–10. https://doi.org/10.1111/conl.12946

Sonter, L. J., Watson, K. B., Wood, S. A., and Ricketts, T. H., 2016. Spatial and temporal dynamics and value of nature-based recreation, estimated via social media. PLoS ONE, 11(9), 1–16. https://doi.org/10.1371/journal.pone.0162372

Souza, C. N., Rodrigues, A. C., Correia, R. A., Normande, I. C., Costa, H. C. M., Guedes-Santos, J., Malhado, A. C. M., Carvalho, A. and Ladle, R. J., 2021. No visit, no interest: How COVID-19 has affected public interest in world's national parks. Biological Conservation, 256, 109015. https://doi.org/10.1016/j.biocon.2021.109015

Spake, R., Bellamy, C., Graham, L., Watts, K., Wilson, T., Norton, L. et al., 2019. A new analytical framework for spatially targeting the management of natural capital. Nature Sustainability, 2(February). https://doi.org/10.1038/s41893-019-0223-4

Sparks, T. H., Dennis, R. L. H., Croxton, P. J., and Cade, M., 2007. Increased migration of Lepidoptera linked to climate change. European Journal of Entomology, 104(1), 139–143. https://doi.org/10.14411/eje.2007.019

Stewart, A. J., Lehmann, N., Corley, I. A., Wang, Y., Chang, Y.-C., Braham, N. A. A. et al., 2023. SSL4EO-L: Datasets and Foundation Models for Landsat Imagery. 1, 1–19. Retrieved from http://arxiv.org/abs/2306.09424

Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10:e13152. https://doi.org/10.7717/peerj.13152

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P. et al., 2019. Machine learning to classify animal species in camera trap images: Applications in ecology. Methods in Ecology and Evolution, 10(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Tate, K. W., Atwill, E. R., Mcdougald, N. K., George, M. R., Tate, K. W., Atwill, E. R. et al., 2017. Society for Range Management Spatial and Temporal Patterns of Cattle Feces Deposition on Rangeland Published by: Society for Range Management Linked references are available on JSTOR for this article: Spatial and temporal patterns of cattle feces deposi. 56(5), 432–438.

TEEB, 2010. Mainstreaming the Economics of Nature: a Synthesis of the Approach, Conclusions and Recommendations of TEEB.

Temple Lang, D., 2020a. RCurl: General Network (HTTP/FTP/…) Client Interface for R. Retrieved from https://cran.r-project.org/package=RCurl

Temple Lang, D., 2020b. XML: Tools for Parsing and Generating XML Within R and S-Plus. Retrieved from https://cran.r-project.org/package=XML

The National River Flow Archive, 2023. https://nrfa.ceh.ac.uk/

The White House, 2022. National Strategy to Develop Statistics for Environmental-Economics Decisions. (January). 89. Retrieved from https://www.whitehouse.gov/wp-content/uploads/2022/08/Natural-Capital-Accounting-Strategy.pdf

Tipping, E., Benham, S., Boyle, J. F., Crow, P., Davies, J., Fischer, U. et al., 2014. Atmospheric deposition of phosphorus to land and freshwater. Environmental Sciences: Processes and Impacts, 16(7), 1608–1617. https://doi.org/10.1039/c3em00641g

Tomlinson, S.J., Carnell, E.J., Dore, A.J., Dragosits, U., 2020. Nitrogen deposition in the UK at 1km resolution, 1990-2017 NERC Environmental Information Data Centre. https://doi.org/10.5285/9b203324-6b37-4e91-b028-e073b197fb9f

UKNEA, 2011. The UK National Ecosystem Assessment: Synthesis of the Key Findings. 1–87. https://doi.org/10.1177/004057368303900411

United Nations, 2020. Population Facts. Retrieved from: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undes_pd_2020_popfacts_urbanization_policies.pdf

United Nations, 2023. Goal 11: Make cities inclusive, safe, resilient and sustainable, Sustainable Development Goals Report, available at: https://www.un.org/sustainabledevelopment/cities/

Van Berkel, D. B., and Verburg, P. H., 2014. Spatial quantification and valuation of cultural ecosystem services in an agricultural landscape. Ecological Indicators, 37, 163–174.

van den Bosch, M., and Ode Sang, A., 2017. Urban natural environments as nature-based solutions for improved public health – A systematic review of reviews. Environmental Research, 158(July), 373–384. https://doi.org/10.1016/j.envres.2017.05.040

Van Zanten, B. T., Van Berkel, D. B., Meentemeyer, R. K., Smith, J. W., Tieskens, K. F., and Verburg, P. H., 2016. Continental-scale quantification of landscape values using social media data. Proceedings of the National Academy of Sciences of the United States of America, 113(46), 12974–12979. https://doi.org/10.1073/pnas.1614158113

Venables, W. N. and Ripley, B. D., 2002. Modern Applied Statistics with S-PLUS. Fourth Edition. Springer.

Villa, F., Bagstad, K. J., Voigt, B., Johnson, G. W., Portela, R., Honzák, M., and Batker, D., 2014. A methodology for adaptable and robust ecosystem services assessment. PLoS ONE, 9(3). https://doi.org/10.1371/journal.pone.0091001

Wade, A. J., Skeffington, R. A., Couture, R. M., Lampa, M. E., Groot, S., Halliday, S. J., 2022. Land Use Change to Reduce Freshwater Nitrogen and Phosphorus Will Be Effective Even with Projected Climate Change. Water (Switzerland), 14(5). https://doi.org/10.3390/w14050829

Wagner, C., Garcia, D., Jadidi, M., and Strohmaier, M., 2015. It's a Man's Wikipedia? Assessing gender inequality in an online encyclopedia. Proceedings of the 9th International Conference on Web and Social Media, ICWSM 2015, 454–463.

Wang, H., Feng, Y., and Ai, L., 2023. Progress of carbon sequestration in urban green space based on bibliometric analysis. Frontiers in Environmental Science, 11. https://doi.org/10.3389/fenvs.2023.1196803

Wang, Z., and Yang, J., 2022. Urbanization strengthens the edge effects on species diversity and composition of woody plants in remnant forests. Forest Ecosystems, 9(January), 100063. https://doi.org/10.1016/j.fecs.2022.100063

Wang, J., Zhou, W., Wang, J., and Qian, Y., 2019. From quantity to quality: enhanced understanding of the changes in urban greenspace. Landscape Ecology, 34(5), 1145–1160. https://doi.org/10.1007/s10980-019-00828-5

Wang, X., Blanchet, F. G., and Koper, N., 2014. Measuring habitat fragmentation: An evaluation of landscape pattern metrics. Methods in Ecology and Evolution, 5(7), 634–646. https://doi.org/10.1111/2041-210X.12198

Watson, R. T., Zakri Director, A., Ramos, M. A., Rosswall, T., Barnard, P., Blasco, D. et al., 2005. Millenium Ecosystem Assessment: Living Beyond Our Means: Natural Assets and Human Well-Being. Retrieved from https://www.millenniumassessment.org/documents/document.429.aspx.pdf

Wearn, O. R., and Glover-Kapfer, P., 2017. Camera-trapping for conservation: a guide to best-practices. WWF Conservation Technology Series 1(1), 181. https://doi.org/10.1002/9781119179313.wbprim0281

Weiskopf, S. R., Rubenstein, M. A., Crozier, L. G., Gaichas, S., Griffis, R., Halofsky, J. E., et al. 2020. Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States. Science of the Total Environment, 733. https://doi.org/10.1016/j.scitotenv.2020.137782

Welsh Government, 2015. Deddf Llesiant Cenedlaethau ' r Dyfodol Well-being of Future Generations (Wales) Act 2015 Well-being of Future Generations (Wales) Act 2015. https://www.gov.wales/well-being-future-generations-act-essentials-html

Wickham, H, 2020. httr: Tools for Working with URLs and HTTP. Retrieved from https://cran.r-project.org/package=httr

Wikimedia Statistics, 2021. Pages to date. Retrieved April 30, 2021, from https://stats.wikimedia.org/#/en.wikipedia.org/content/pages-to-date/normal%7Ctable%7Call%7Cpage_type~content%7Cmonthly

Wilkins, E. J., Wood, S. A., and Smith, J. W., 2021. Uses and Limitations of Social Media to Inform Visitor Use Management in Parks and Protected Areas: A Systematic Review. Environmental Management, 67(1), 120–132.

Willcock, S., Martínez-López, J., Hooftman, D. A. P., Bagstad, K. J., Balbi, S., Marzo, A. et al., 2018. Machine learning for ecosystem services. Ecosystem Services, 33, 165–174. https://doi.org/10.1016/j.ecoser.2018.04.004

Willi, M., Pitman, R.T., Cardoso, A.W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., and Fortson L., 2018. Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution 10: 80-91.

Withers, P. J. A., and Jarvie, H. P., 2008. Delivery and cycling of phosphorus in rivers: a review. The Science of the Total Environment, 400(1–3), 379–395. https://doi.org/10.1016/j.scitotenv.2008.08.002

Withers, P. J. A., Neal, C., Jarvie, H. P., and Doody, D. G., 2014. Agriculture and eutrophication: Where do we go from here? Sustainability (Switzerland), 6(9), 5853–5875. https://doi.org/10.3390/su6095853

Wolch, J. R., Byrne, J., and Newell, J. P., 2014. Urban green space, public health, and environmental justice: The challenge of making cities "just green enough." Landscape and Urban Planning, 125, 234–244. https://doi.org/10.1016/j.landurbplan.2014.01.017

Wong, Y., and Rosindell, J., 2022. Dynamic visualisation of million-tip trees: The OneZoom project. Methods in Ecology and Evolution, 13(2), 303–313. https://doi.org/https://doi.org/10.1111/2041-210X.13766

Wood, S. A., Guerry, A. D., Silver, J. M., and Lacayo, M., 2013. Using social media to quantify nature-based tourism and recreation. Scientific Reports, 3. https://doi.org/10.1038/srep02976

Wood, S. A., Winder, S. G., Lia, E H., White, E. M., Crowley, C. S. L., and Milnor, A. A., 2020. Next-generation visitation models using social media to estimate recreation on public lands. Scientific Reports, 10(1), 1–12. https://doi.org/10.1038/s41598-020-70829-x

World Resources Institute, 2014. Global Forest Watch. Retrieved July 1, 2019, from www.globalforestwatch.org

Wu, J., Shen, W., Sun, W., and Tueller, P. T., 2002. Empirical patterns of the effects of changing scale on landscape metrics. Landscape Ecology, 17(8), 761–782. https://doi.org/10.1023/A:1022995922992

Wu, X., Zhu, X., Wu, G.-Q., and Ding, W., 2014. Data Mining with Big Data. Proceedings of 2017 11th International Conference on Intelligent Systems and Control, ISCO 2017, 26(1), 246–250. https://doi.org/10.1109/ISCO.2017.7855990

Yang, C., He, X., Wang, R., Yan, F., Yu, L., Bu, K., Yang, J., Chang, L., Zhang, S., 2017. The effect of urban green spaces on the urban thermal environment and its seasonal variations. Forests 8:153

Yang, Y., Jia, Y., Ling, S., and Yao, C., 2021. Urban natural resource accounting based on the system of environmental economic accounting in Northwest China: A case study of Xi'an. Ecosystem Services, 47(February 2020), 101233. https://doi.org/10.1016/j.ecoser.2020.101233

Yu, Q., Ji, W., Prihodko, L., Ross, C. W., Anchang, J. Y., and Hanan, N. P., 2021. Study becomes insight: Ecological learning from machine learning. Methods in Ecology and Evolution, 12(11), 2117–2128. https://doi.org/10.1111/2041-210X.13686

Zhang, D, 2020. rsq: R-Squared and Related Measures. Retrieved from https://cran.r-project.org/package=rsq

Zhou, W., Wang, J., and Cadenasso, M. L., 2017. Effects of the spatial configuration of trees on urban heat mitigation: A comparative study. Remote Sensing of Environment, 195, 1–12. https://doi.org/10.1016/j.rse.2017.03.043

Zhu, Z., and Woodcock, C. E., 2014. Remote Sensing of Environment Continuous change detection and classi fi cation of land cover using all available Landsat data. Remote Sensing of Environment, 144, 152–171. https://doi.org/10.1016/j.rse.2014.01.011

Zimmermann, E., Bracalenti, L., Piacentini, R., and Inostroza, L., 2016. Urban Flood Risk Reduction by Increasing Green Areas for Adaptation to Climate Change. Procedia Engineering, 161, 2241–2246. https://doi.org/10.1016/j.proeng.2016.08.822

Zou, K. H., O'Malley, A. J., and Mauri, L., 2007. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, *115*(5), 654–657. https://doi.org/10.1161/CIRCULATIONAHA.105.594929

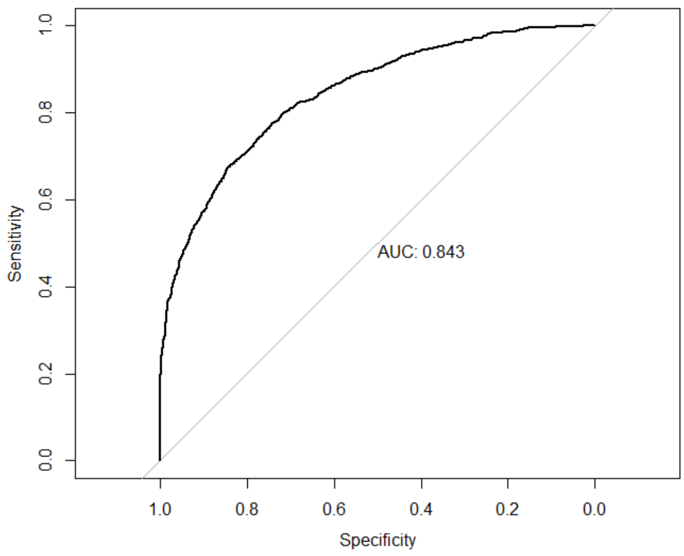
Zurlini, G., Petrosillo, I., Aretano, R., Castorini, I., Darpa, S., De Marco, A., Pasimeni, M.R., Semeraro, T., Zaccarelli, N., 2014. Key fundamental aspects for mapping and assessing ecosystem services: predictability of ecosystem services providers at scales from local to global. Annali di Botanica 4:53.

Zuur, A. F., Ieno, E. N., Walker, N. J., Saveliev, A. A., and Smith, G. M., 2009. Mixed Effects Models and Extensions in Ecology with R. In Statistics for Biology and Health (pp. 1–574.). New York, NY, USA: Springer.

Appendices

Appendix 1

(a)



(b)

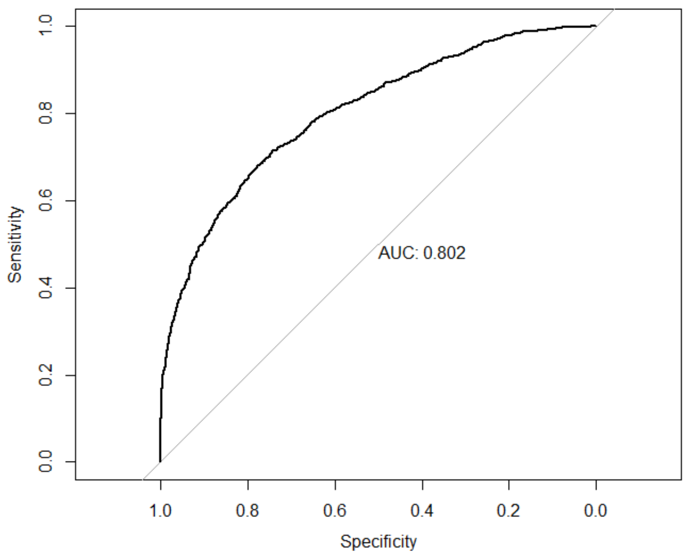
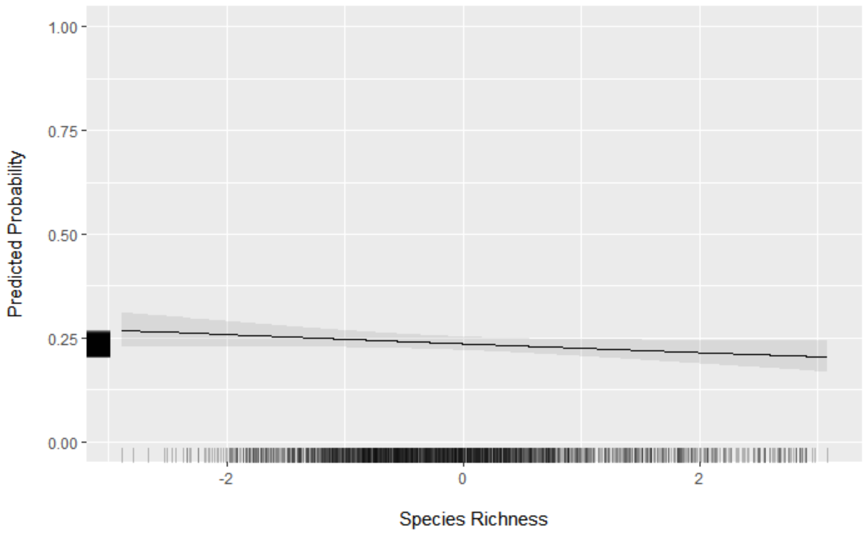


Figure A1.1: ROC curves for (a) the best model for the Flickr data and (b) the best model for the Wikipedia data. The plots show the sensitivity and specificity of the model at different probability cutoffs. The AUC is an overall summary of diagnostic accuracy and a good model will have a high AUC. AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy (Zou et al. 2007).

Appendix 2



(b)

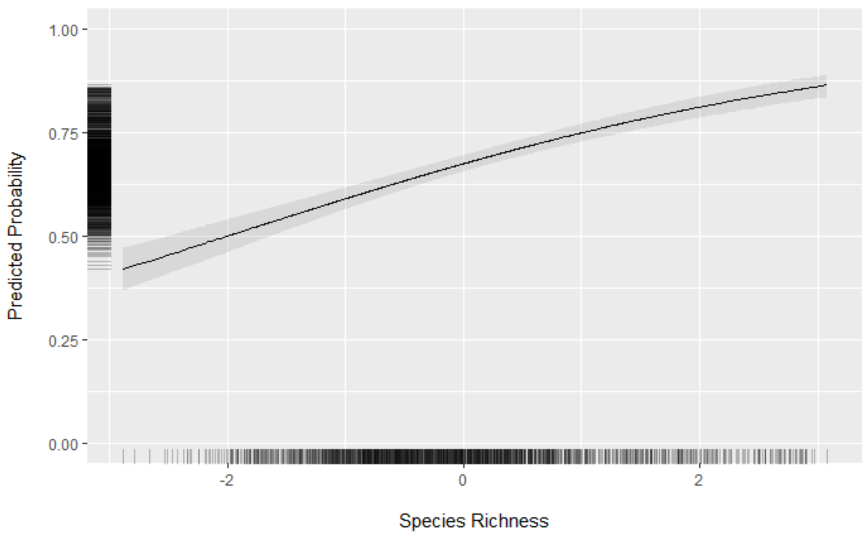


Figure A2.1: Results of the models for (a) the Flickr data and (b) the Wikipedia data. The plots show the predicted probability of finding at least one Flickr photo (a) or Wikipedia page (b) against species richness in designated areas without a coastal location or waterbody, while other variables are kept at their mean. Species richness was centred around the mean and scaled by its SD. Ticks on the plot margins represent the data, the lines represent the predictions from the model and the lighter shaded areas are the 95% confidence intervals.

Appendix 3

This section briefly explains the additional analysis carried out in order to ensure that the very small amount of spatial autocorrelation found in the residuals of the two binomial glm models is not influencing the conclusions drawn.

Overall, there is no clear spatial autocorrelation visible in the residuals of the two glm models for the Flickr data (Figure A3.1) and the Wikipedia data (Figure A3.2). However, Moran's I for the residuals shows a very small amount of spatial autocorrelation, that is nonetheless significant, for both the model of the Flickr data (observed = 0.03, expected = -0.0002, p-value < 0.001) and the model of the Wikipedia data (observed = 0.04, expected = -0.0002, p-value < 0.001). The semivariograms shows that the spatial autocorrelation is no longer an issue at 4000 m, as the semivariograms levels off (Figure A3.3).

To ensure that our conclusions are not affected by this small amount of spatial autocorrelation we repeatedly took a random sub-sample of the designated areas, ensuring that those included were more than 4000 m apart. Each sub-sample included about 2400 designated areas (from the total of 6349), and these were used to run binomial glms for the Flickr and Wikipedia data. The coefficients for the covariates in the models varied by less than 0.1 compared to the glm using all of the data, meaning that our conclusions remain unchanged.

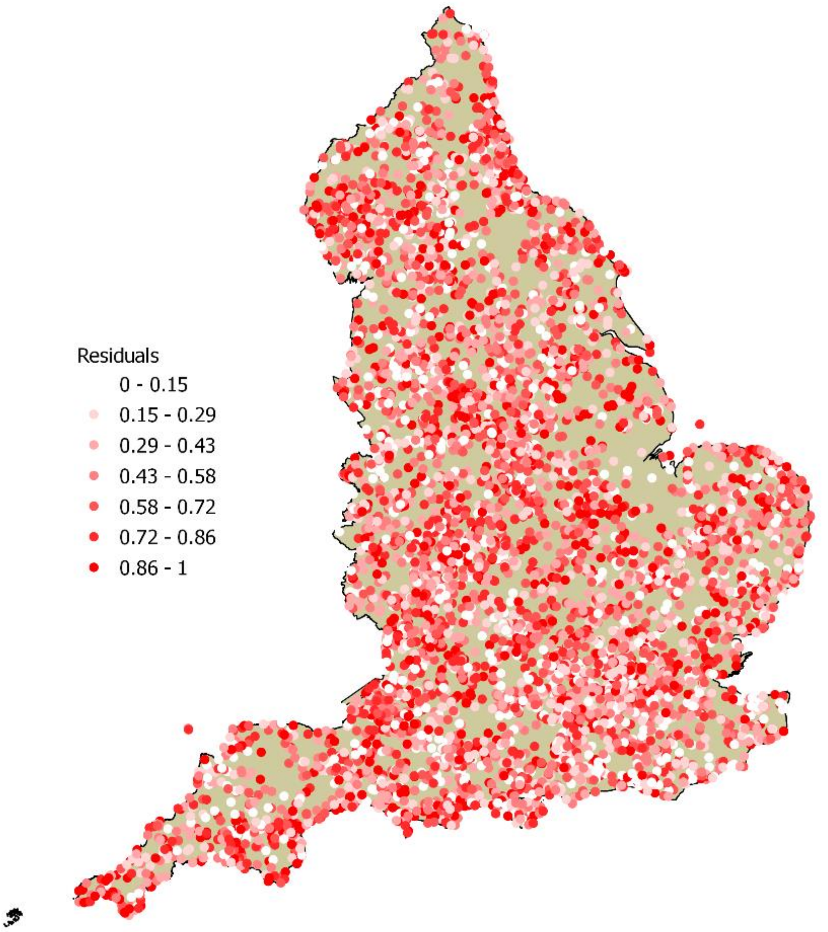


Figure A3.1: Map of the simulated (DHARMa) residuals from the model of the Flickr data. Each point represents the centroid of a designated area. The range of DHARMa residuals is 0 to 1, which makes them easier to visualise and interpret than the "raw" residuals of the model.

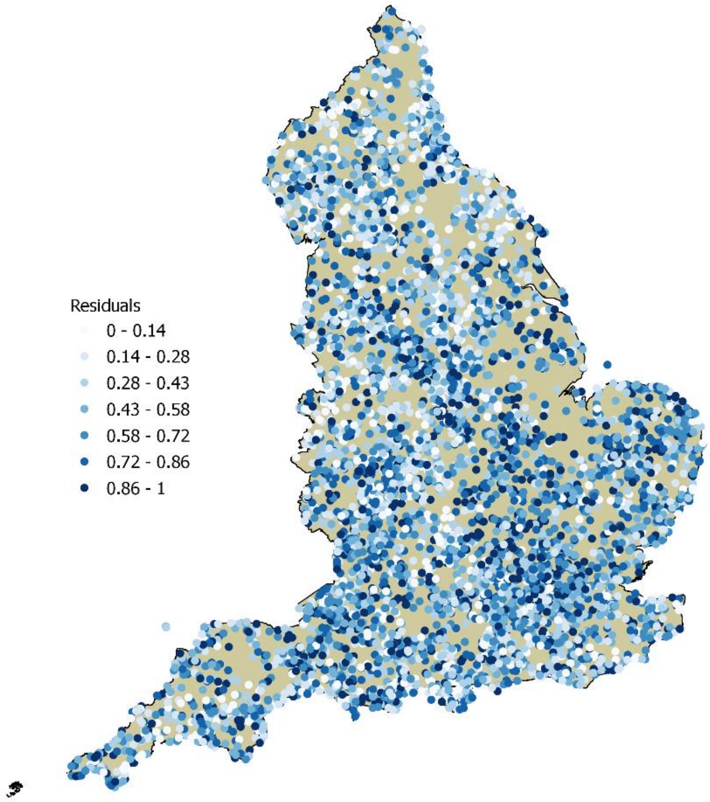
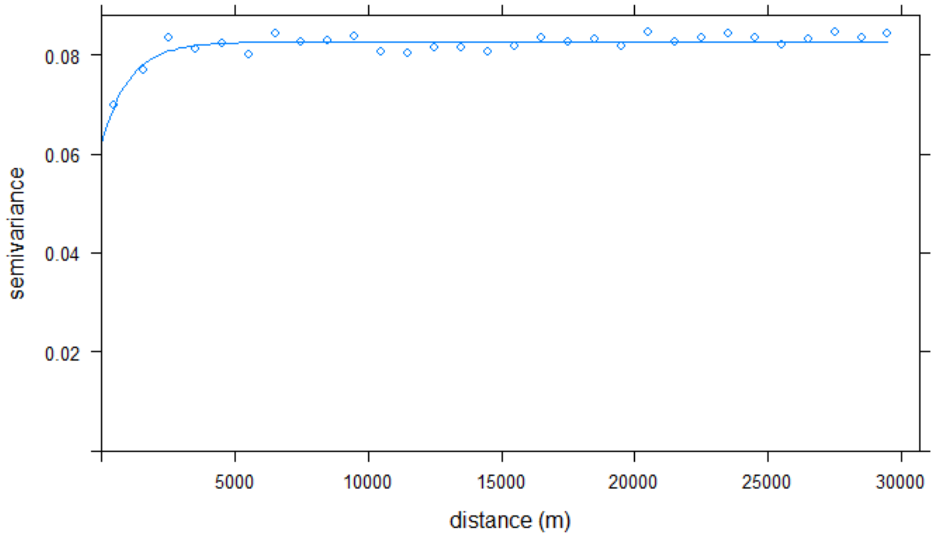


Figure A3.2: Map of the simulated (DHARMa) residuals from the best model of the Wikipedia data. Each point represents the centroid of a designated area. The range of DHARMa residuals is 0 to 1, which makes them easier to visualise and interpret than the "raw" residuals of the model.

(a)



(b)

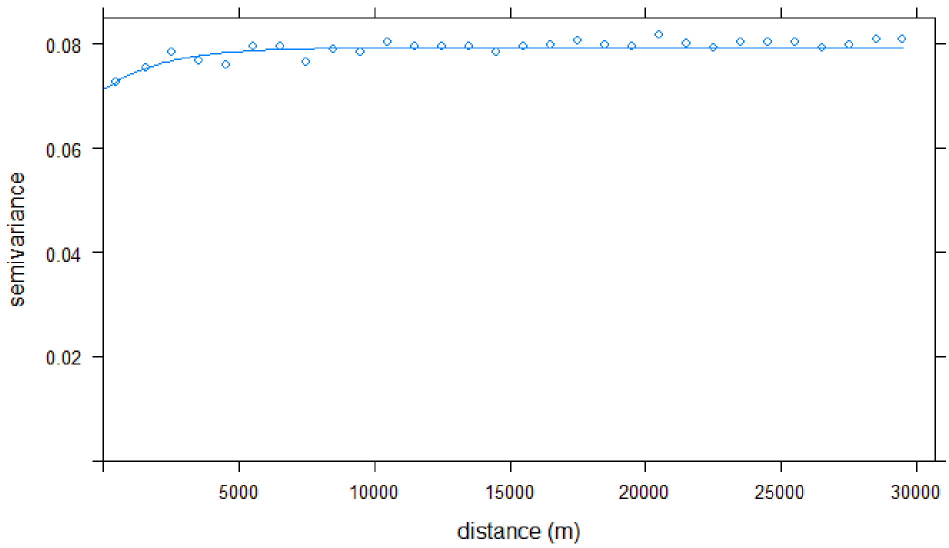


Figure A3.3: Semivariogram for the simulated (DHARMa) residuals of the best model for the Flickr data (a) and the Wikipedia data (b).

Appendix 4

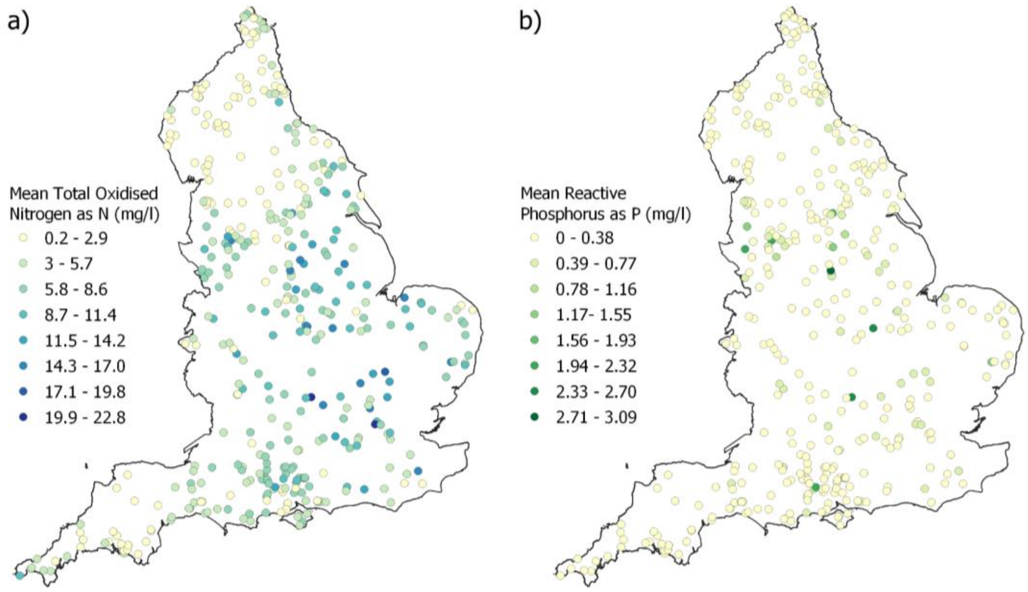


Figure A4.1: Mean Total Oxidised Nitrogen (a) and mean Reactive Phosphorus (b) for the years 2015-2019 at the monitoring stations initially selected based on minimum data availability, and after ensuring that the monitoring stations' catchments did not overlap, as described in Section 3.2.1.2. For Total Oxidised Nitrogen n = 404, for Reactive Phosphorus n = 383. We included a subset of this data in the study, choosing those with high and low population density (those below the first quartile and above the fourth quartile, respectively), as described in Section 3.2.2.

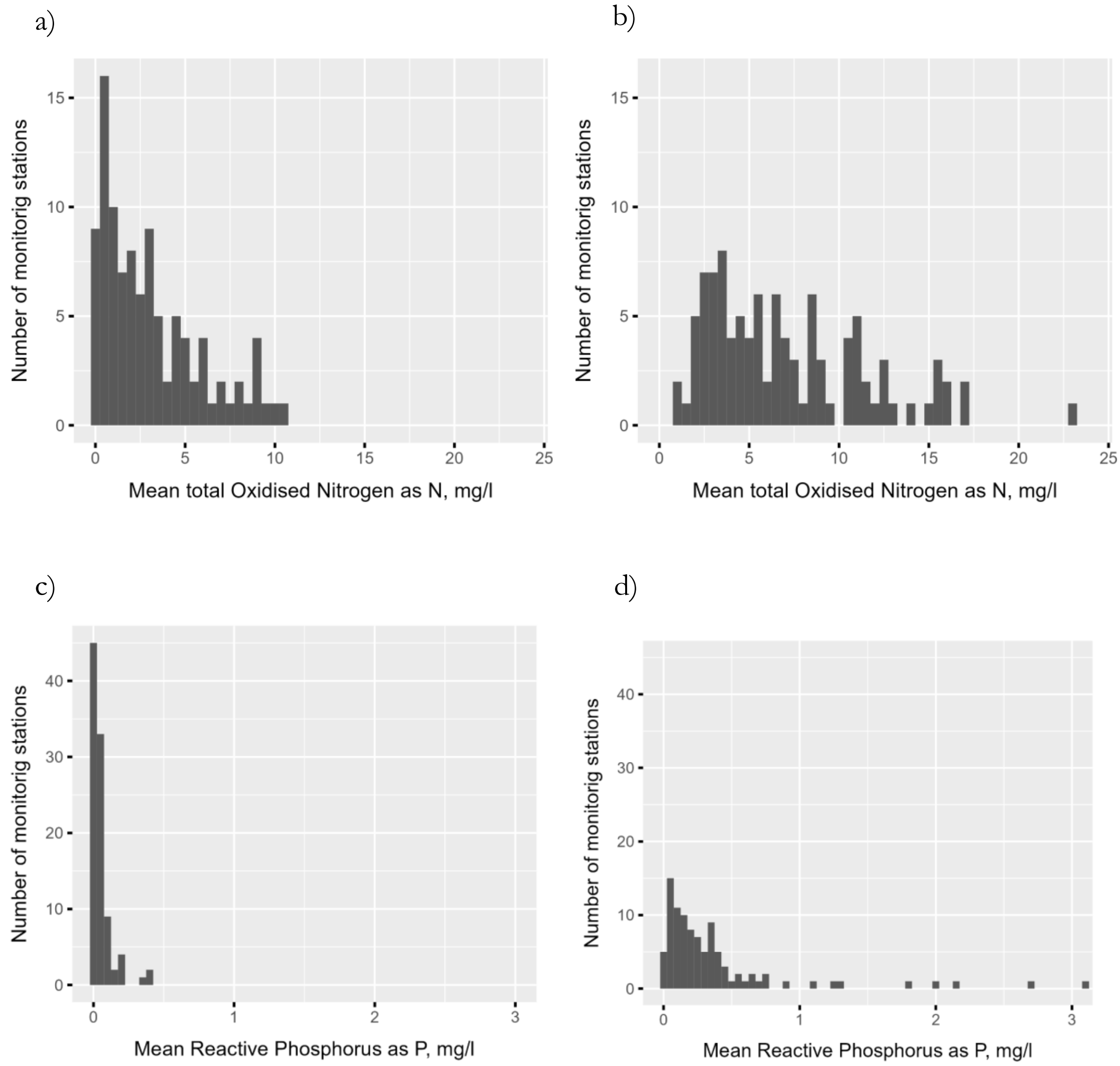


Figure A4.2: Distribution of the dependent variable of the four datasets used in this study, that is the mean Total Oxidised Nitrogen (TON) for catchments with low population density (a) and high population density (b) (n = 101 in each case), and mean Reactive Phosphorus (RP) for catchments with low population density (c) and high population density (d) (n = 96 in each case). The mean of TON and RP was taken over the years 2015 to 2019.

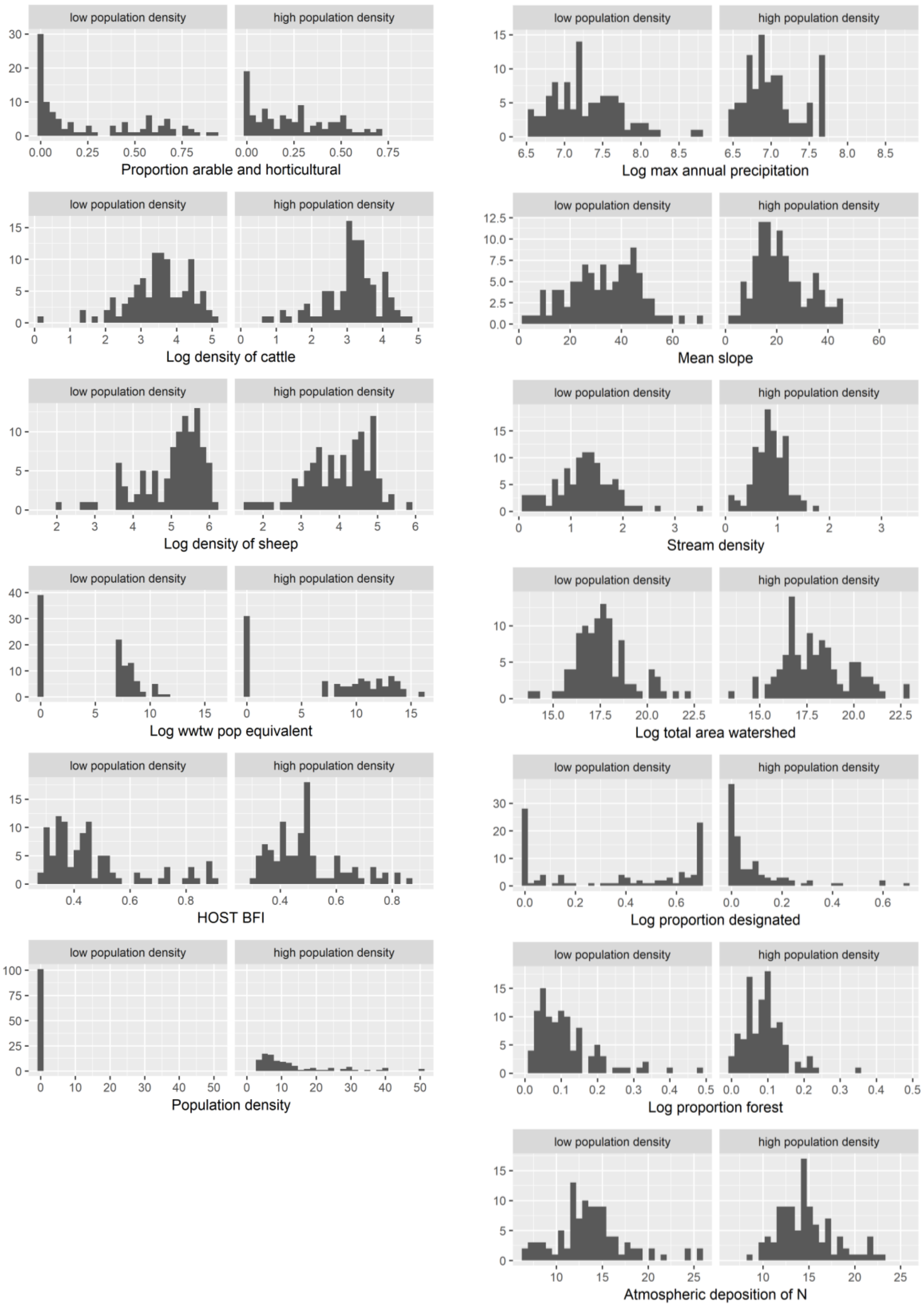


Figure A4.3: Distribution of the independent variables for the models of TON (model of catchments with low population density and model of catchments with high population density).

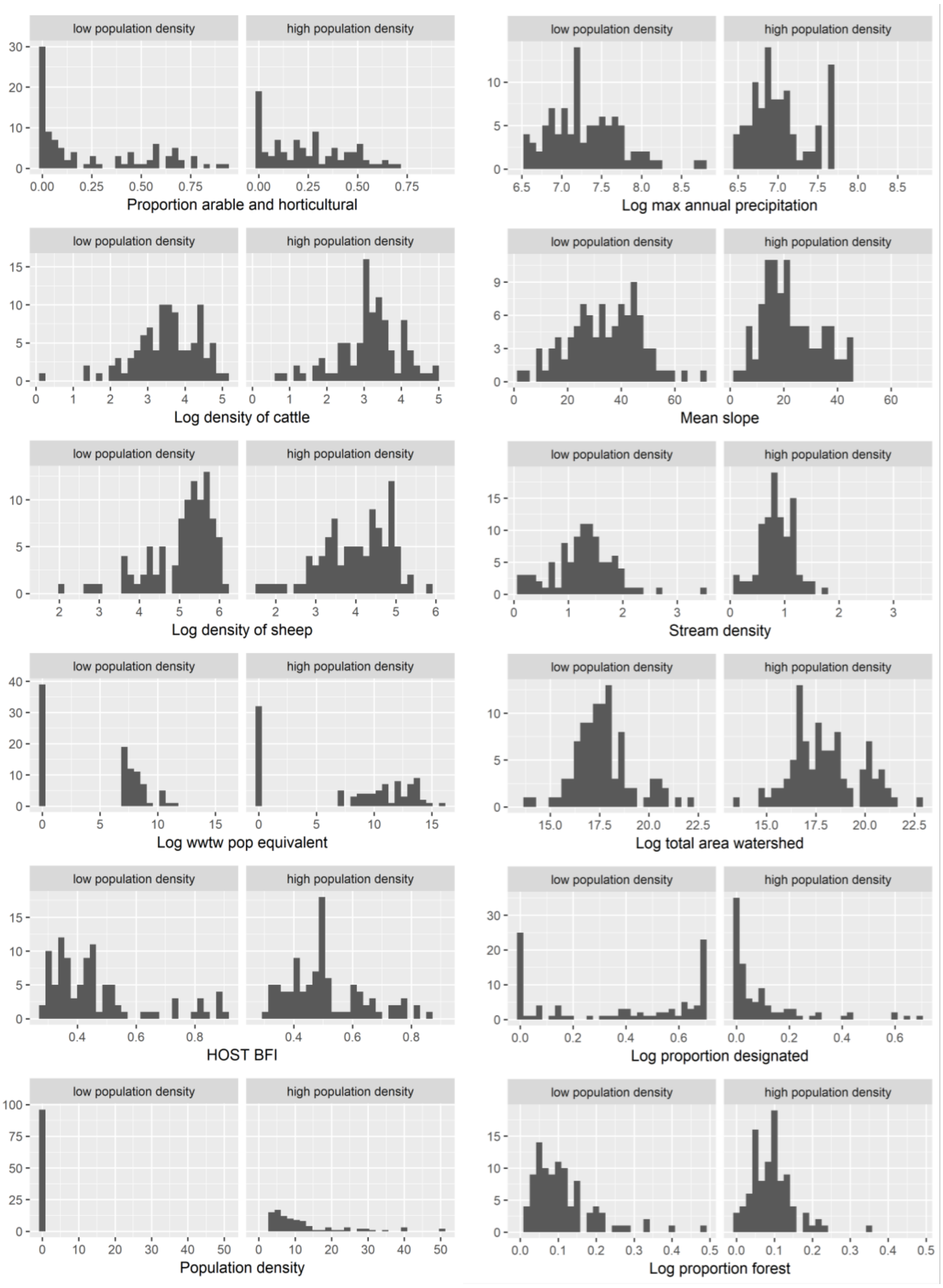


Figure A4.4: Distribution of the independent variables for the models of RP (model of catchments with low population density and model of catchments with high population density). Note that the distribution of the independent variables for the models of RP are almost identical to those for TON (Figure A4.3) because most monitoring stations appear in both datasets.

Appendix 5

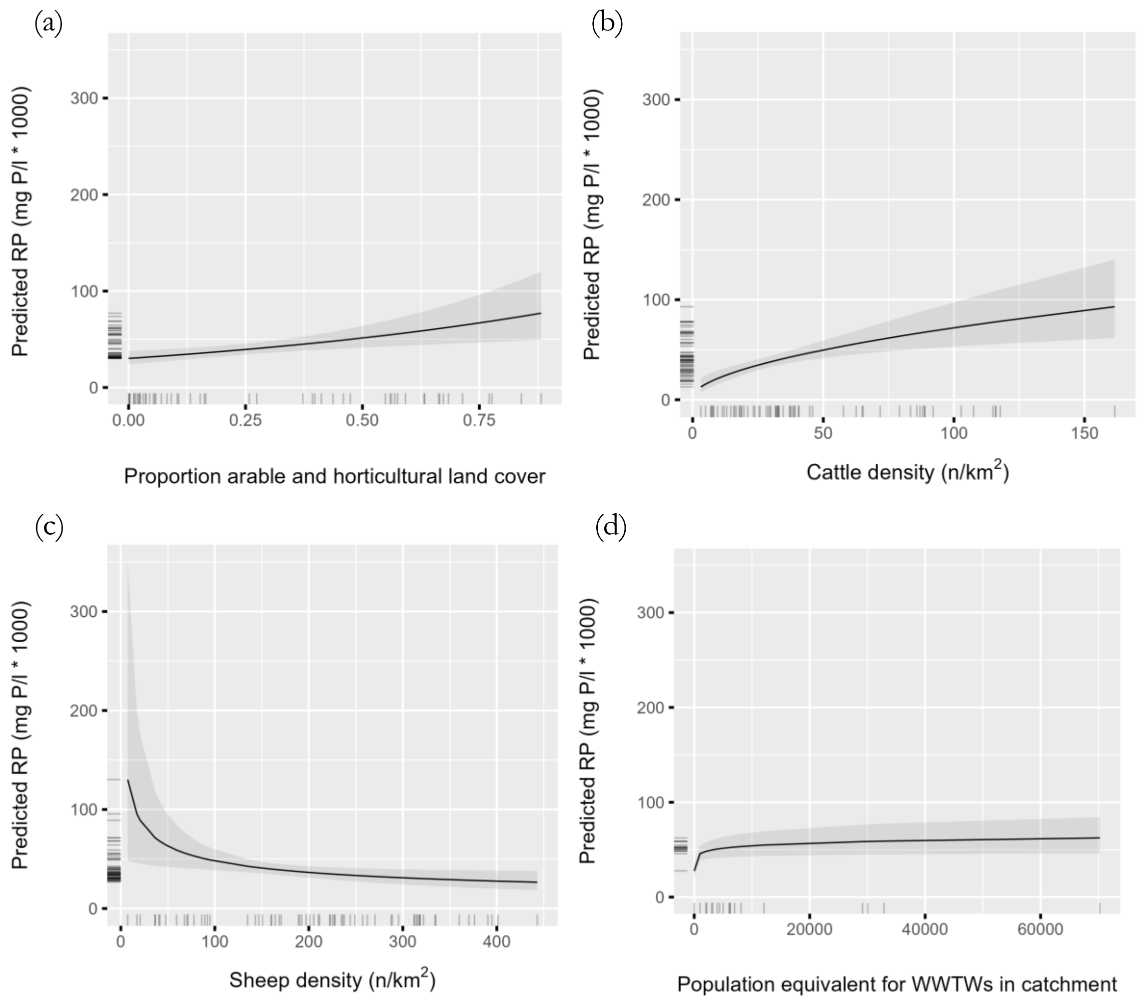


Figure A5.1: Effect plots for the generalised linear model of RP in catchments with low population density. The plots show the predicted RP concentration*1000 against the proportion of arable and horticultural land cover (a), cattle density (b), sheep density (c) and the population equivalent of the WWTWs in the catchment (d), with all other variables kept at their mean. The population equivalent of the WWTWs in the catchment, cattle density and sheep density were exponentiated before plotting, as the predictor in the model was the log of the population equivalent of WWTWs, log of cattle density and log of sheep density.

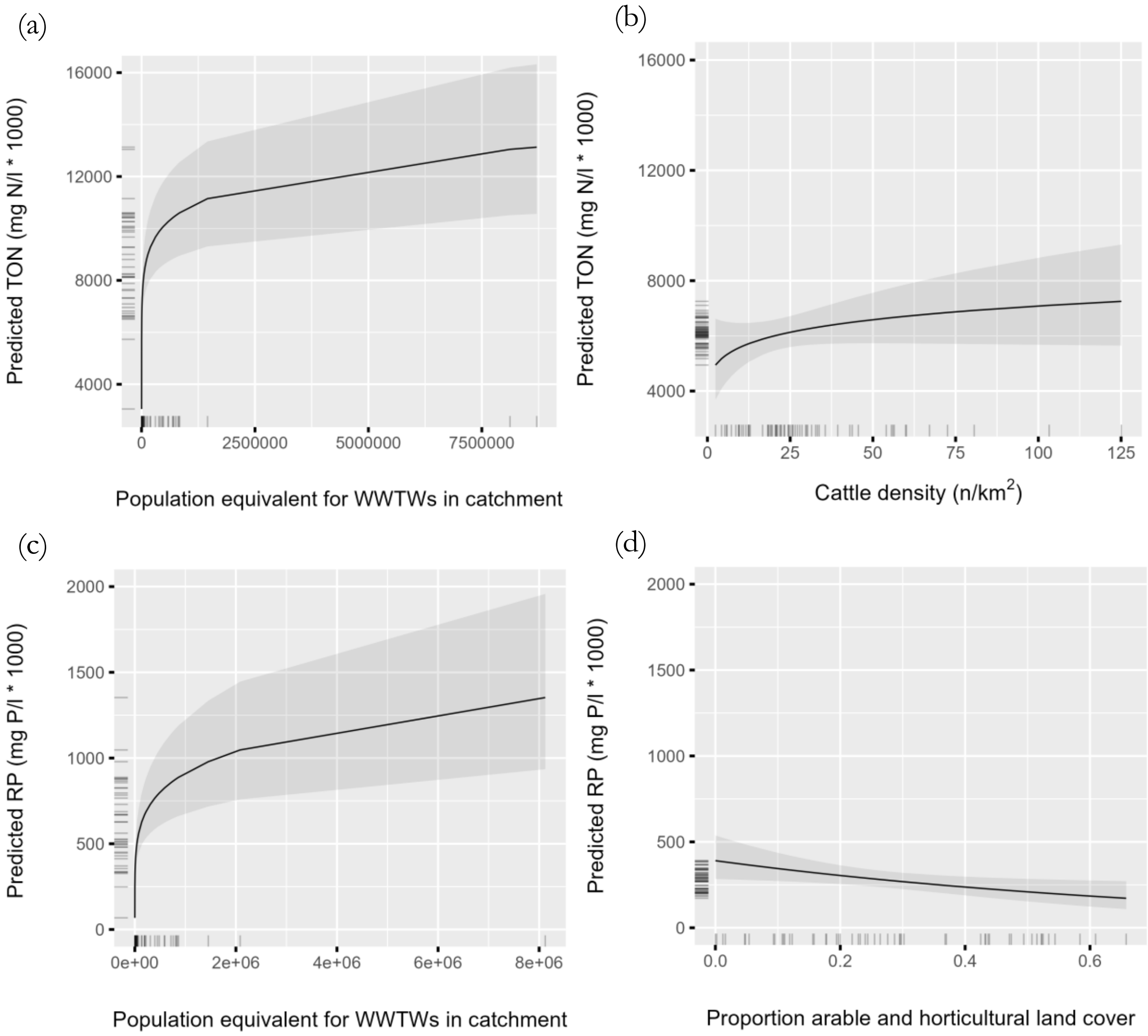


Figure A5.2: Effect plots for the negative binomial generalised linear model of TON in catchments for catchments with high population density (a, b) and for the negative binomial generalised linear model of RP in catchments for catchments with high population density (c, d). The plots show the predicted TON and RP concentration*1000 against the population equivalent of the WWTWs in the catchment (a and c), the predicted TON against cattle density in the catchment (b) and the predicted RP against the proportion of arable and horticultural land cover in the catchments (c), with all other variables kept at their mean. The population equivalent of the WWTWs in the catchment and the cattle density were exponentiated before plotting, as the predictor in the model was the log of the population equivalent of WWTWs and the log of cattle density.

Appendix 6

Table A6.1: *Land use classes and corresponding resistance values used by the Omniscape algorithm (in the form of a habitat suitability map). The underlying land use map of Greater London includes data from a London-wide database of open spaces (Greenspace Information for Greater London CIC 2023c) and the OS MasterMap Greenspace Layer (Ordnance Survey Limited 2022). The resistance values were chosen based on the example in the Omniscape documentation, local knowledge of Greater London and various trial runs.*

Land use	Resistance
Allotments, Community Growing Spaces or City Farms	5
Amenity - Residential or Business	50
Amenity - Transport	50
Bowling Green	50
Cemetery	5
Camping or Caravan Park	50
Golf Course	50
Institutional Grounds	50
Land Use Changing	50
Natural	1
Other Sports Facility	50
Play Space	50
Playing Field	50
Private Garden	5
Public Park or Garden	5
Religious Grounds	5
School Grounds	50
Tennis Court	50
Civic Spaces	50
Green Corridors	1
Natural and Semi-natural Urban Greenspace	1
Other	50
Other Urban Fringe	5
Buildings and large rivers	Absolute barrier

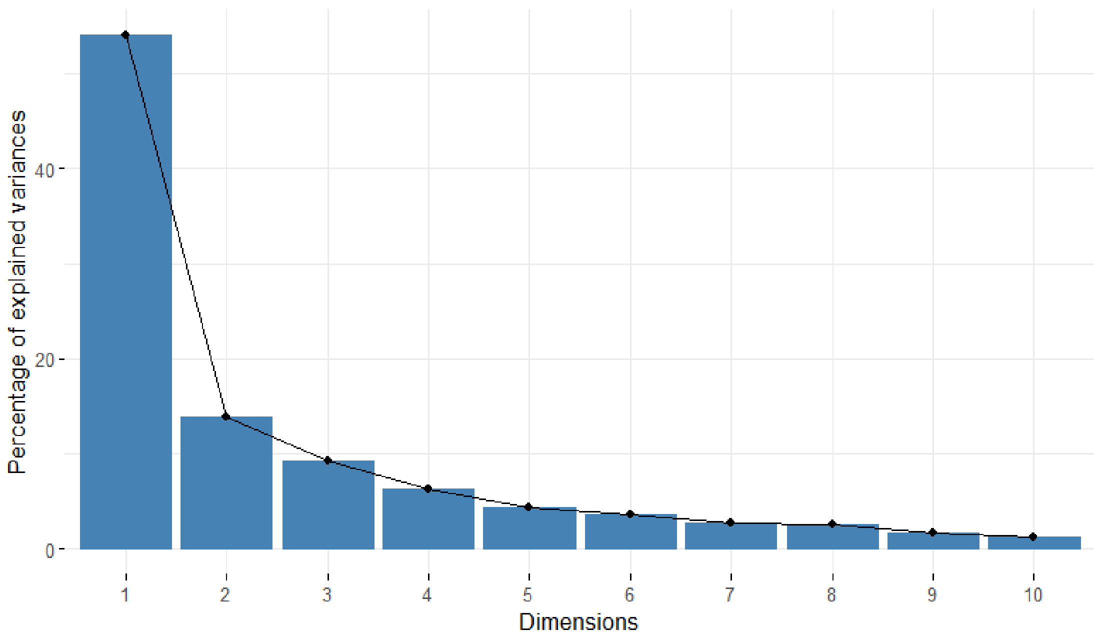
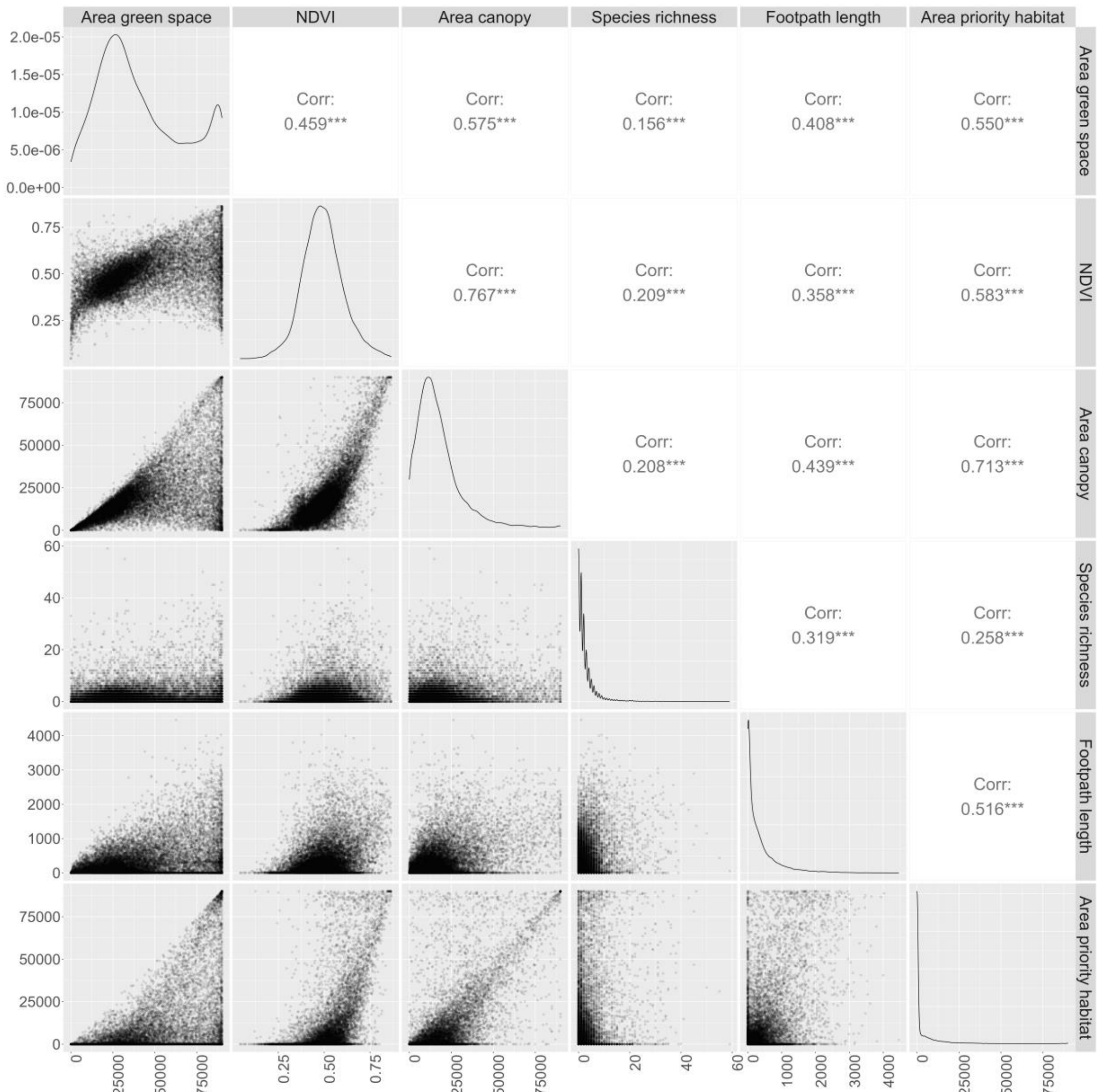


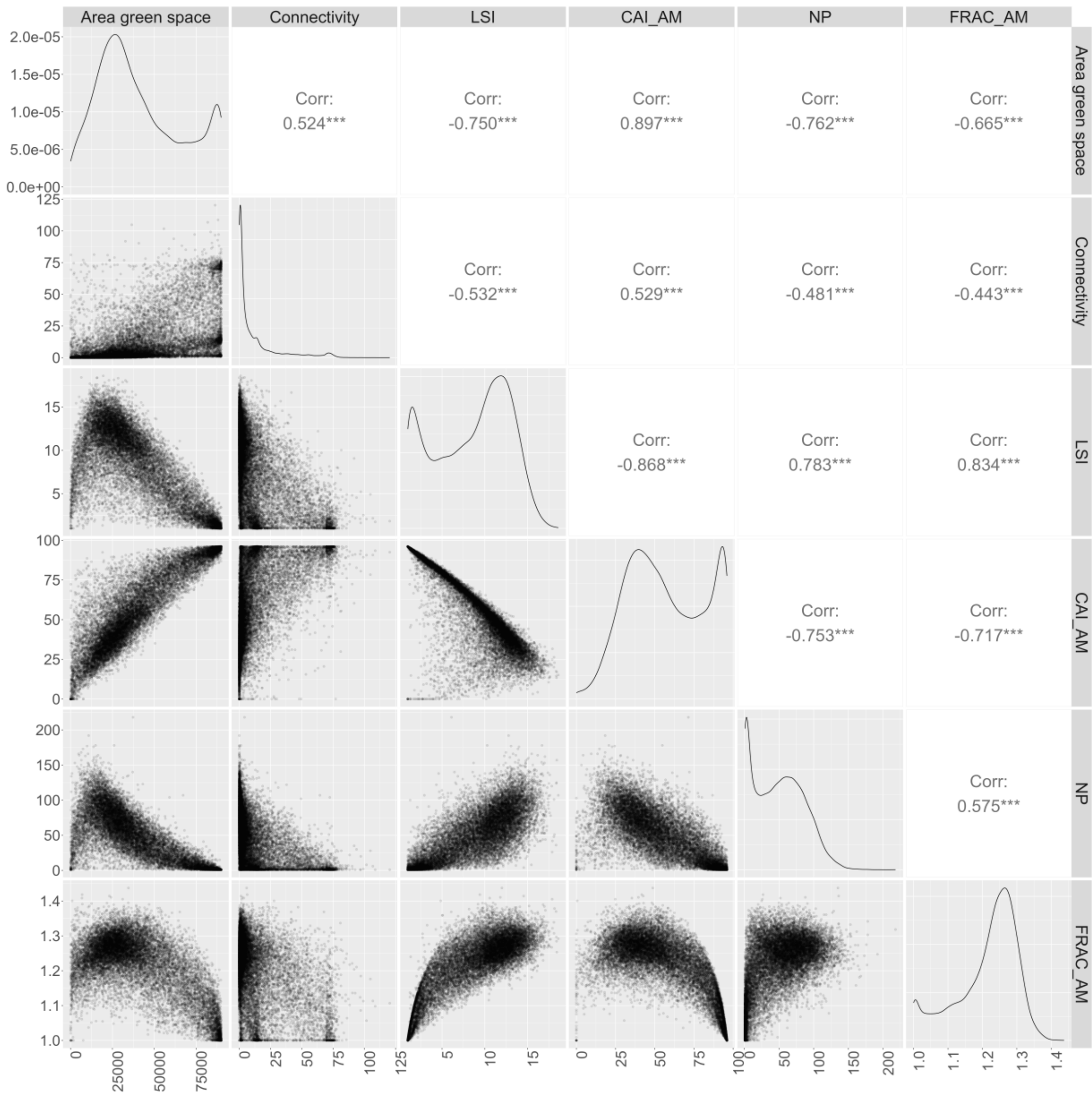
Figure A6.1: Scree plot of the explained variance, with each dimension representing a principal component.

(a)



Area green space
NDVI
Area canopy
Species richness
Footpath length
Area priority habitat
Corr: 0.459***
Corr: 0.575***
Corr: 0.156***
Corr: 0.408***
Corr: 0.550***
Corr: 0.767***
Corr: 0.209***
Corr: 0.358***
Corr: 0.583***
Corr: 0.208***
Corr: 0.439***
Corr: 0.713***
Corr: 0.319***
Corr: 0.258***
Corr: 0.516***

(b)



Area green space
Connectivity
LSI
CAI_AM
NP
FRAC_AM
Corr:
0.524***
Corr:
-0.750***
Corr:
0.897***
Corr:
-0.762***
Corr:
-0.665***
Corr:
-0.532***
Corr:
0.529***
Corr:
-0.481***
Corr:
-0.443***
Corr:
-0.868***
Corr:
0.783***
Corr:
0.834***
Corr:
-0.753***
Corr:
-0.717***
Corr:
0.575***

(c)

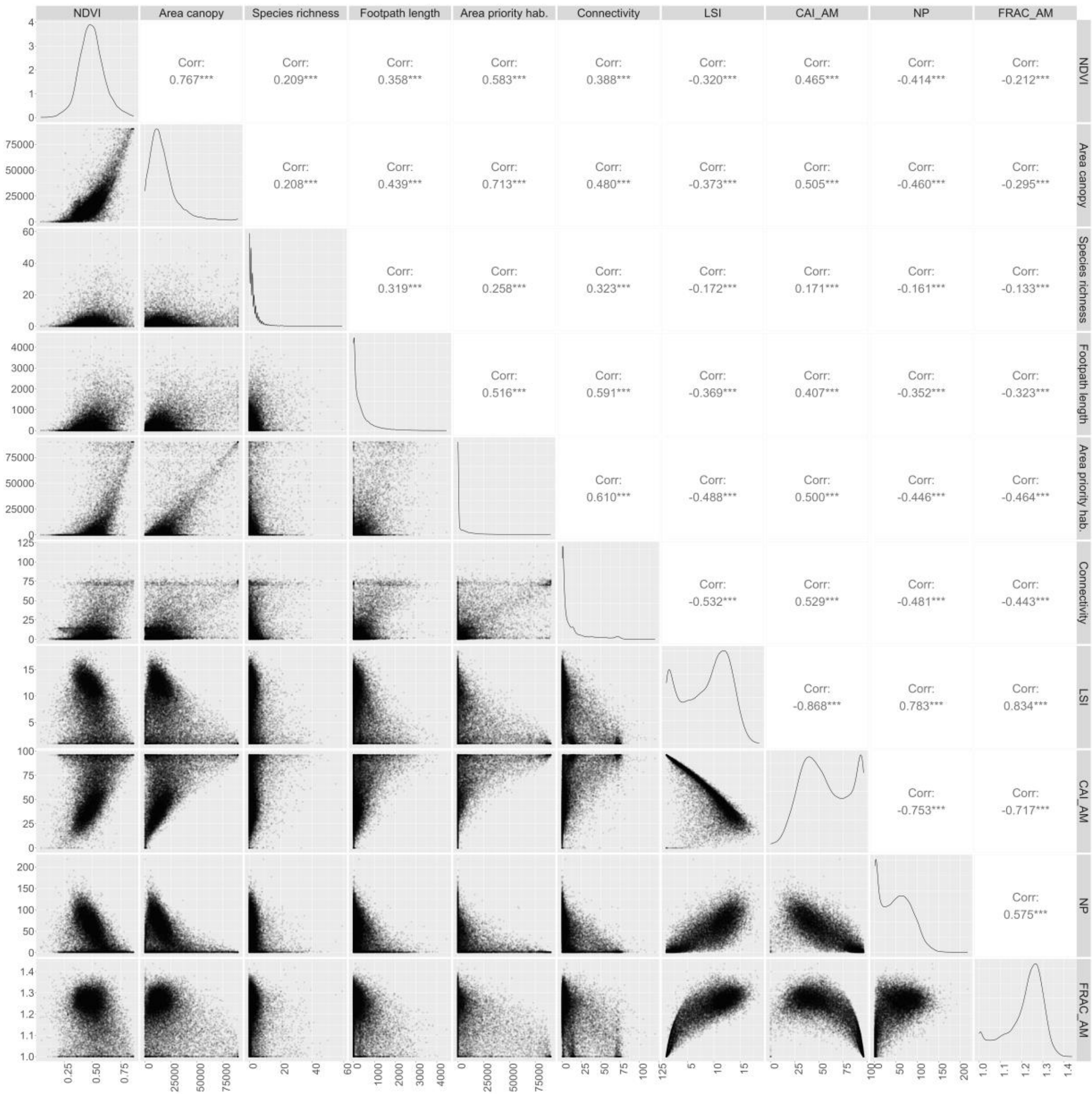


Figure A6.2: Pearson correlation coefficients, density plots and scatter plots for the indicators used in this study, showing the correlation between the indicators for quantity and quality (a), quantity and spatial configuration (b) and quality and spatial configuration (c). The correlation coefficients were all statistically significant with $p < 0.001$.

Appendix 7

Table A7.1: *Loadings for the first (PC1) and second (PC2) principal components from the indicators for quantity, quality and spatial configuration for Greater London calculated for a 1200 m grid (a) and a 5000 m grid (b), showing the contribution of each variable. The signs of the loadings are arbitrary, and randomly assigned, so may differ between different programs for principal component analysis (R Core Team 2022).*

(a)

Dimension	Indicator	PC1*	PC2*
Quantity	Area green space	-0.34*	0.21
Quality	Species richness	-0.20*	-0.39
Quality	NDVI	-0.26*	-0.19
Quality	Total area of tree canopy cover	-0.29*	-0.20
Quality	Total length of footpaths	-0.22*	-0.52*
Quality	Total area of habitat of principal importance	-0.31*	-0.26
Spatial configuration	CAI_AM	-0.35*	0.26
Spatial configuration	LSI	0.33*	-0.30
Spatial configuration	LPI	-0.34*	0.28
Spatial configuration	NP	0.32*	-0.29
Spatial configuration	Landscape connectivity	-0.29*	-0.26

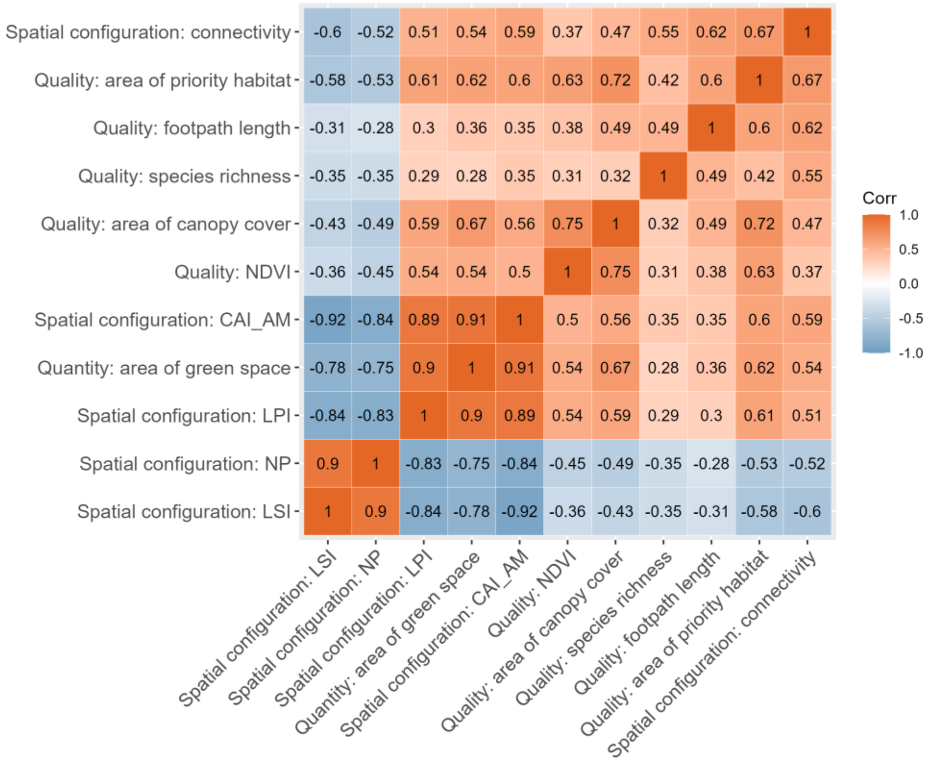
*Indicates statistically significant ($p < 0.05$)

(b)

Dimension	Indicator	PC1*	PC2*
Quantity	Area green space	-0.33*	0.15
Quality	Species richness	-0.17*	0.27
Quality	NDVI	-0.29*	0.07
Quality	Total area of tree canopy cover	-0.26*	0.37
Quality	Total length of footpaths	-0.12*	0.56*
Quality	Total area of habitat of principal importance	-0.33*	0.30
Spatial configuration	CAI_AM	-0.37*	-0.26
Spatial configuration	LSI	0.34*	0.35
Spatial configuration	LPI	-0.37*	-0.19
Spatial configuration	NP	0.33*	0.35
Spatial configuration	Landscape connectivity	-0.29*	0.07

*Indicates statistically significant ($p < 0.05$)

(a) 1200 m grid



(b) 5000 m grid

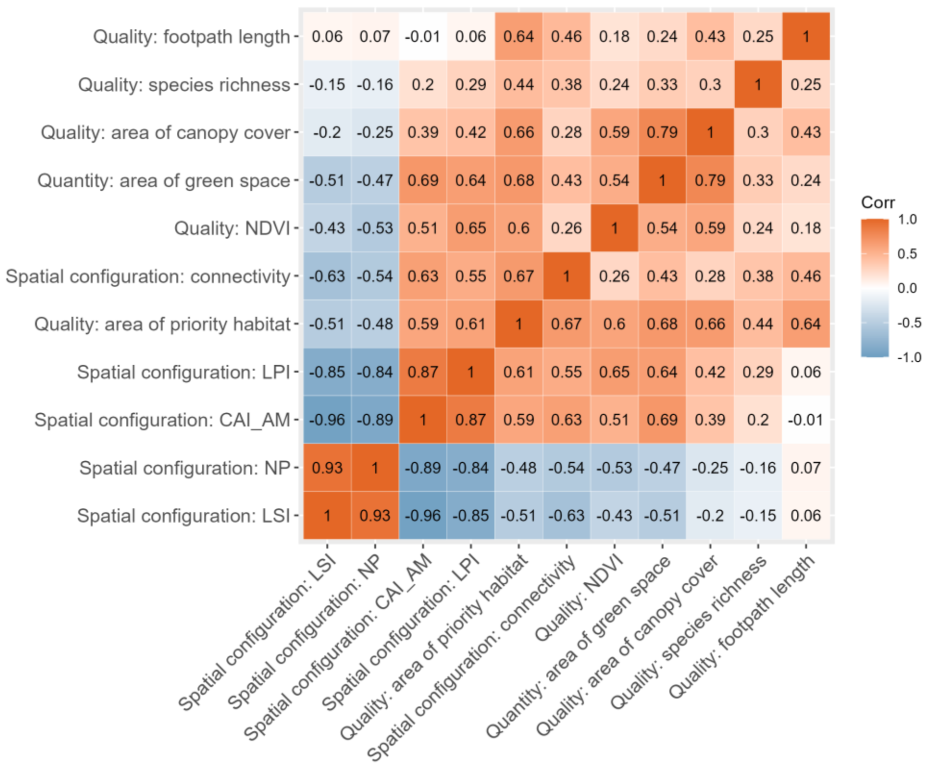


Figure A7.1: Pearson correlation coefficients for the indicators used in this study, calculated for a 1200 m grid (a) and a 5000 m grid (b). The correlation coefficients were all statistically significant with $p < 0.001$.