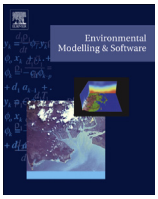
Environmental Modelling & Software 68 (2015) 219–232

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft



A geospatial framework to support integrated biogeochemical modelling in the United Kingdom

S. Greene [a], [*], P.J. Johnes [b], J.P. Bloomfield [c], S.M. Reaney [d], R. Lawley [e], Y. Elkhatib [f], J. Freer [b], N. Odoni [b], C.J.A. Macleod [g], B. Percy [h]

[a] *Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, United Kingdom*
[b] *School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1SS, United Kingdom*
[c] *British Geological Survey, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, United Kingdom*
[d] *Department of Geography, Durham University, Durham, DH1 3LE, United Kingdom*
[e] *British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, NG12 5GG, United Kingdom*
[f] *School of Computing & Communications, Lancaster University, LA1 4WA, Lancaster, United Kingdom*
[g] *The James Hutton Institute, Craigiebuckler, Aberdeen, AB15 8QH, United Kingdom*
[h] *Department of Meteorology, University of Reading, Whiteknights, Reading, RG6 6AL, United Kingdom*

ARTICLE INFO

Article history:
Received 9 December 2014
Accepted 16 February 2015
Available online

Keywords:
Water resource management
Spatial targeting
Pollution
Cloud computing
Decision support

ABSTRACT

Anthropogenic impacts on the aquatic environment, especially in the context of nutrients, provide a major challenge for water resource management. The heterogeneous nature of policy relevant management units (e.g. catchments), in terms of environmental controls on nutrient source and transport, leads to the need for holistic management. However, current strategies are limited by current understanding and knowledge that is transferable between spatial scales and landscape typologies. This study presents a spatially-explicit framework to support the modelling of nutrients from land to water, encompassing environmental and spatial complexities. The framework recognises nine homogeneous landscape units, distinct in terms of sensitivity of nutrient losses to waterbodies. The functionality of the framework is demonstrated by supporting an exemplar nutrient model, applied within the Environmental Virtual Observatory pilot (EVOp) cloud cyber-infrastructure. We demonstrate scope for the use of the framework as a management decision support tool and for further development of integrated biogeochemical modelling.

Software availability

Title: Geospatial framework to support integrated biogeochemical modelling in the United Kingdom
Developers: Greene, S.[a], Johnes, P.J., Bloomfield, J.P., Reaney, S.M., Lawley, R., Elkhatib, Y., Freer, J., Odoni, N., MacLeod, C.J.A., Percy, B.
Contact corresponding author: Address: [a]Centre for Ecology & Hydrology, Maclean Building, Benson Lane, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, United Kingdom
Email: shgreene@tcd.ie
Tel: +44 (0) 1491 692495
Fax: +44 (0) 1491 692424
First available: 2012
Hardware: Any GIS enabled device
Software: GIS software
Format: ESRI shapefile (vector polygon)
Size: 585 MB
Availability: Contact author directly at shgreene@tcd.ie

1. Introduction

Nutrient enrichment of inland and coastal waters and the subsequent decline in their ecological quality present a widespread problem for water resource management (Sutton et al., 2011; McGonigle et al., 2012; Liu et al., 2012). Effective management for preventative and remedial action at the catchment, regional or landscape scale requires explicit consideration of the complexities of the hydrological and biogeochemical controls that act in unity to

---

[*] Corresponding author. Tel.: +44 (0) 1491 692495; fax: +44 (0) 1491 692424.
E-mail address: shgreene@tcd.ie (S. Greene).

http://dx.doi.org/10.1016/j.envsoft.2015.02.012

deliver nutrients to waterbodies (Pärn et al., 2012; Greene et al., 2013; Robson, 2014). However, the development of such integrated management strategies is limited by the availability of science understanding and knowledge that is transferable between spatial scales and landscape typologies.

Observational data used for informing process understanding of nutrient transfer is often generated at relatively small experimental scales. Findings from these studies are not necessarily directly transferable for application at the whole catchment scale without the development of a modelling solution, bespoke for the system and the data available to drive that model (Haygarth et al., 2012; Ye et al., 2012). As catchment area increases, considerable variation in the hydrological and biogeochemical controls on nutrients transpire, reflecting variations in upstream catchment characteristics such as geology, soils, land use and topography. A modelling solution is also often based on the determination of partial fractions of the total nutrient load exported from land to water, underestimating the scale and impact of the enrichment problem, and is therefore likely to misinform the direction of environmental management aimed at mitigating the problem (Burt and Johnes, 1997; Johnes, 2007; Yates and Johnes, 2013).

Policy makers and environmental managers seeking to develop management and mitigation options for impacted systems have had to either (1) fund the development of site-specific modelling and monitoring programmes for each catchment of interest, (2) use knowledge acquired from inadequate, low resolution or partial monitoring of a limited range of nutrient fractions in the catchment of interest or neighbouring catchment, or (3) use knowledge acquired from high resolution studies in systems which might not be directly comparable with the catchment of interest. In order to deliver effective comprehension of nutrient cycling and export dynamics under current and potential environmental change conditions there is an urgent need to develop better mechanisms for the transfer of knowledge and science understanding between data-rich and data-poor systems (Beven and Alcock, 2012).

A range of approaches have been developed to tackle the need for better understanding and management of nutrients in the environment. Simpler correlative statistical modelling approaches, such as the Global News model (Seitzinger et al., 2010), can generate visually attractive simulations of catchment behaviour at regional to global scales, but lack a physical basis (or representation of the specific physical conditions controlling functional behaviour in differing environments). Such approaches are often inaccurate and uncertain when reduced to a scale suitable for environmental management, generating high risk when used to support operational management and policy development. Dynamic process-based modelling approaches with a physical basis, such as the Integrated Catchment (INCA) modelling suite for nitrogen (N), phosphorus (P), carbon (C) and sediment (Whitehead et al., 1998; Wade et al., 2002; Futter et al., 2007; Lazar et al., 2010), provide the opportunity to capture the science understanding generated by high resolution research on catchment behaviours at a range of scales. However expert knowledge and high concomitant costs are currently associated with the calibration of the model(s) to local conditions in each application (Dean et al., 2009).

Less data-intensive models that represent intrinsic nutrient retention and release capacity as a function of environmental attributes provide a suitable alternative approach. Such models have been developed in a number of countries, and based on measureable properties of the landscape, they provide optimised parameter ranges for key drivers of catchment function, related to regional landscape typologies. These include an explicit representation of landscape form and function either within optimised parameter ranges, or embedded within a regionalised framework. Examples include the export coefficient model (ECM) applied within the context of a geoclimatic region framework (Johnes et al., 1996, 2007; Johnes and Butterfield, 2002), the integrated models IMAGE and INTEGRATOR, the Indicator Database for European Agriculture (IDEAg) covering the agriculture sector and sewage systems, the Emission Database for Global Atmospheric Research (EDGAR) covering atmospheric emissions from all sectors, and the Unified EMEP model calculating atmospheric transport and deposition models applied across the whole of Europe in the most recent European Nitrogen Assessment (Leip et al., 2011). These models tend to operate at a coarser temporal scale in order to fit the model predictions to the average annual behaviour of a catchment, but are spatially explicit in simulating typical catchment behaviours within similar, quasi-homogenous region types. As these are based on measureable properties of the landscape, they offer opportunities for aggregating data and knowledge from plot to catchment, regional and global scale, and the transfer of knowledge from data-rich to data-poor environments by using local spatially-explicit high resolution data from the environment.

This paper aims to present an improved regionalised framework that explicitly integrates the myriad of environmental attributes controlling nutrient retention and loading exported to waterbodies across the United Kingdom (UK). Two objectives are set. First, an earlier framework by Johnes et al. (2007), used as a foundation, has been extensively refined in an iterative way to improve the region classification methodology, spatial resolution of source datasets and geographic extent. Second, we demonstrate how this framework has been applied at catchment and broader spatial scales by supporting an exemplar nutrient model. Instead of traditional off-line processing, the framework and associated model has been hosted by a novel cloud cyber-infrastructure, developed in the Environmental Virtual Observatory Pilot (EVOp) project, funded by the Natural Environment Research Council (NERC). This setup provides efficient processing time together with a user-friendly geospatial web portal (Elkhatib et al., 2013; Emmett et al., 2014; Vitolo et al., 2015). A discussion of how such a framework and approach will contribute to advancing nutrient modelling and management is also included.

2. Materials and methods

The functionality of the foundation geoclimatic region framework (Johnes et al., 2007) was demonstrated using the ECM developed by Johnes (1996). For context we give a brief description of the ECM and application to the original framework, including evaluation metrics and limitations (section 2.1). The development of the new framework is described, together with details on source datasets, classification rules and evaluation and application metrics (section 2.2). Application of the framework supporting the exemplar ECM within the cloud computing architecture and evaluation of model estimates is also outlined (section 2.3). All development and evaluation metrics are guided by recommendations by Bennett et al. (2013).

2.1. Background

2.1.1. Export coefficient model

The ECM developed by Johnes (1996) is based on the semi-distributed approach that calculates the nutrient load exported to any water body (freshwater, estuarine or marine) as the sum of the total nutrient load exported to that water body from all sources within its catchment. Different coefficients are adopted for individual crops, as well as for types of livestock units and the coefficients for humans are determined based on the discharge and treatment efficiency of domestic sewage in large Wastewater Treatment Works (WwTW), small packet Sewage Treatment Works (STW) and septic tank systems. The model also takes into account nitrogen fixation by plants (varying by crop type) and deposition of both N and P from atmospheric sources. The ECM is outlined as:

$$L = \sum_{i=1}^{n} E_i(A_i(I_i)) + p \tag{1}$$

where L is the load of nutrients (TN or TP), E_i is the export coefficient for nutrient source i, A_i is the area of the catchment occupied by land use type i, or number of livestock type i, or people, I_i is the input of nutrients to source i, and p is the total input of nutrients from atmospheric deposition. Rates of nutrient input and rates of nutrient export are based on spatially explicit monitoring and experimental data for each major livestock type and land cover category, and vary between differing geoclimatic region types.

2.1.2. Foundation framework

2.1.2.1. Iterative calibration and performance evaluation. The ECM was simplified to run for a limited number of landscape unit types in England and Wales that share similar functional behaviour in terms of process controls on TN and TP cycling and loading to waterbodies (Johnes, 1996, 2007; Johnes and Butterfield, 2002). These were originally defined based on the major classes of geoclimatic region identified in the 1st Land Utilisation Survey of Britain (1931–1940) and are illustrated in Fig. 1 (a). Geoclimatic regions represent landscape units comprising broadly similar climate, geology, soil types, topography and natural vegetation cover which have, therefore, similar ranges of nutrient export potential (and nutrient retention capacity) as a function of flow volume, timing and routing from land to stream. Generic sets of export coefficients (unit-specific parameter values) were derived for each of these geoclimatic regions and land use class. The coefficients were selected from plot scale experiments and small headwater catchment scale research, and were optimised to reflect the intrinsic nutrient retention capacity of each region type. Using a quantitative approach by coupling real and modelled values (Bennett et al., 2013), the selected coefficients were calibrated in an initial application to observed TN and TP data for one water year in one catchment for which high resolution observational data were available (data-rich). The calibrated coefficients selected for each nutrient source in each region were then tested in a rigorous multiple testing procedure, with the calibrated model run using the same coefficients for a series of time slices from the historic records for that catchment. If the model predicted within ±10% of observed loads the model coefficients were accepted as tested and validated for that catchment and were then applied using the same internally tested coefficient set to at least two other catchments with 10 years of observational data in that geoclimatic region type. If the model failed the initial internal testing, it was re-calibrated until the test threshold was met. If the model failed the second, external test when run for comparable catchments in the same geoclimatic region, model calibration was revisited.

The catchments in which the model was tested against observed TN and TP data are shown in Fig. 1 (b), coded to indicate the dominant original geoclimatic region type for each catchment. Modelled data were also plotted against corresponding high resolution observational data from 75 data-rich catchments. A minimum of three catchments for each of the six geoclimatic region types were included. Close fits ($R^2 = 0.98$ for both N and P for 75 catchments and >150 pairs of data). Further details on the calibration and test steps are given in publications by Johnes (1996, 1999), Johnes et al. (1996, 2007) and Johnes and Butterfield (2002). Model fit with reconstruction of past P loading on two lakes in England using diatom indices was also confirmed and tested against observed P concentration data for the lakes (Bennion et al., 2005).

2.1.2.2. Limitations. The original framework provided the first opportunity for integrated national scale TN and TP modelling. However, the framework was limited by the lack of available (at that time) high resolution environmental information with which to derive the geoclimatic regions mapping, and was furthermore limited to the geographical extent of England and Wales. The general approach has been revisited in the current research to generate an improved framework, at a finer spatial resolution, allowing for better discrimination of nutrient export behaviours in small, data-poor catchments and across a wider range of landscape types, including extension to cover Scotland and Northern Ireland. The new classified regions include lowland peatland, coastal plain and estuarine environments that had previously provided the least precise estimations when lumped with coarser geoclimatic region classes.

2.2. Refined framework

2.2.1. Iterative derivation of explicit homogeneous landscape units

An iterative approach was adopted in developing and testing the classification of distinct geoclimatic regions. The methodological approach involved the first comprehensive use of the British Geological Survey (BGS) Parent Material Model (PMM) data, which was provided by the BGS at a spatial resolution of 1:50 000 (Lawley, 2011). These source data provide geospatial information on the basic foundations of soils, their structure, drainage and geochemistry across the UK. Metrics of slope gradient (degrees) extracted from a 1:50 000 resolution digital elevation model (DEM) and 1 km^2 gridded mean annual runoff data (mm day^{-1}) were also used as additional source data for the regionalised classification. ArcGIS Desktop 10.0 was used for data manipulation.

Using current knowledge and expert opinion from a multidisciplinary working group of biogeochemists, hydrologists, geologists and soil scientists, a rule-based methodology (Fig. 2) was developed to separate out nine homogenous geoclimatic unit types (Table 1, Fig. 3). These rules were based on the characteristics of the three sets of source data used, in terms of landscape sensitivities to TN and TP export and retention. First, the characteristics of the original six geoclimatic regions (pink, green, blue, brown, yellow and red) were reviewed and explicit rules for separation better defined (e.g. unambiguous geology, slope and runoff thresholds). A further three regions (olive, orange and purple) were identified as key landscape types that were not properly represented in the original classification. Using data based on their biogeochemical and hydrological characteristics these three landscape types were extracted to form distinct units.

The major landscape unit defined for the upland regions in the UK is classified as yellow (in the web version) in Fig. 3. This represents those areas of the UK with the highest potential rates of N and P flow from land to water due to steep slopes, high rainfall and low baseflow index. They have the lowest actual (realised) rates of export, however, as these upland moorland peat areas (>300 m above mean sea level (amsl)) mainly support forestry, low intensity sheep production, and to a lesser

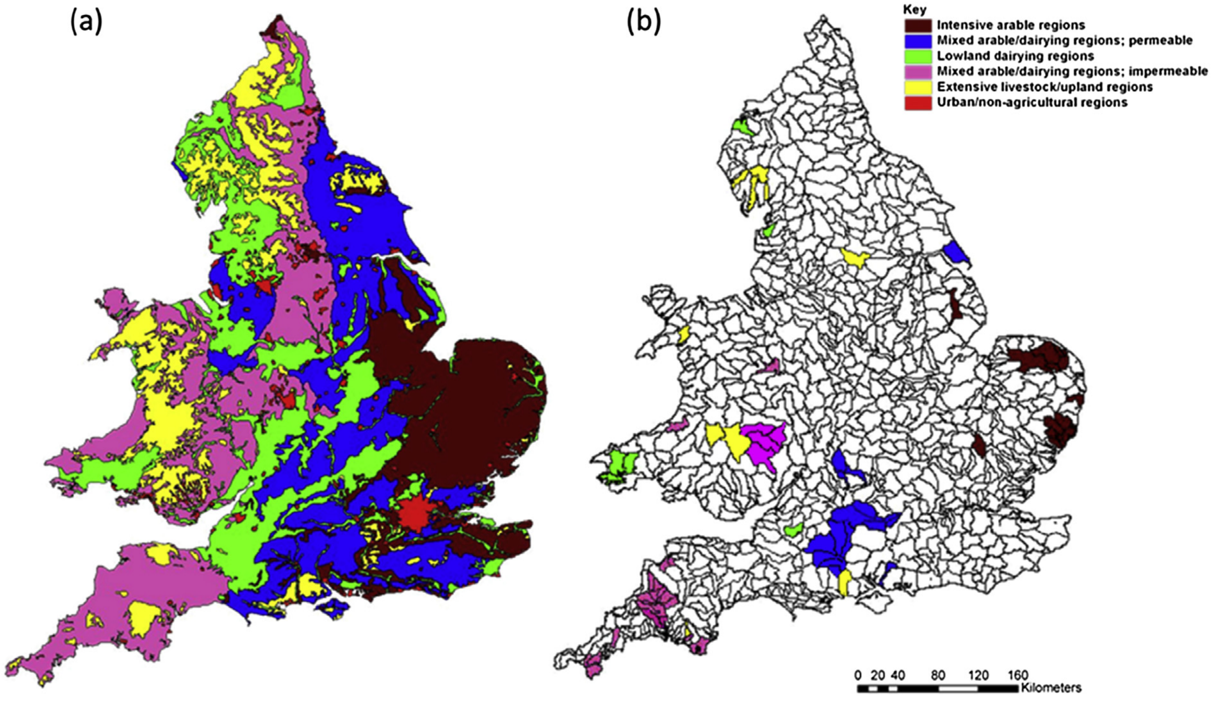
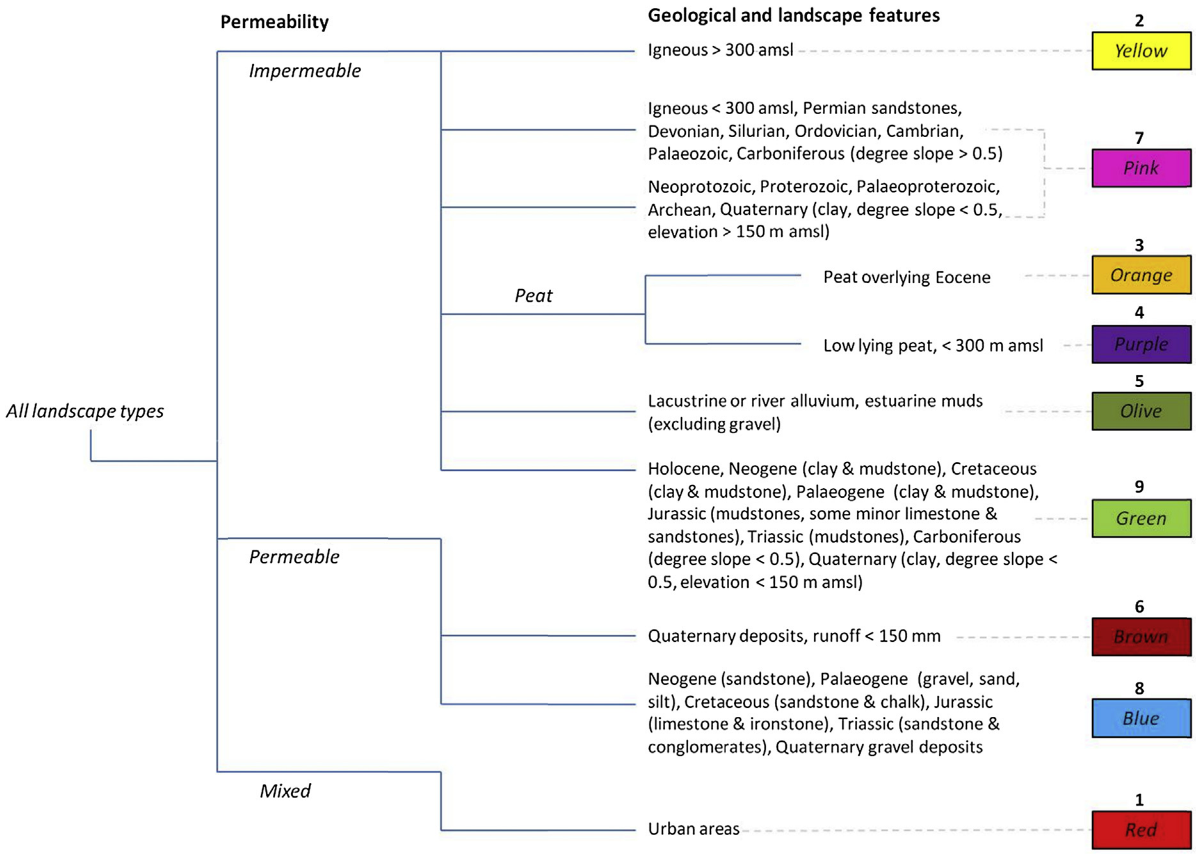


Fig. 1. (a) The original geoclimatic region framework for England and Wales (after Johnes et al., 1996, 2007; Johnes and Butterfield, 2002). (b) A minimum of three test catchments with a minimum of 10 years observed nutrient loading data were used for testing modelled estimates with observed data for each geoclimatic region.



Geology and soil data derived from British Geological Survey Parent Material Model (50 m resolution). Elevation (above mean sea level (amsl)) and slope data were derived from a 50m digital elevation model. Mean annual runoff data (1961-2008) were extracted at 1 km resolution. The numbers over the coloured geoclimatic region classes indicate the sequence the class was extracted from the data (e.g. *Pink (7)* was extracted from the data remaining after *Red (1)* to *Brown(6)* were classified).

Fig. 2. Categorisation tree showing the data selection process leading to the classification of the nine geoclimatic regions.

Table 1
Characteristics and percentage of the nine revised geoclimatic region types in the UK.

Geoclimatic region	Characteristics	% coverage
Blue	Moderately sloping, hilly, overlying permeable sedimentary bedrock. Free draining.	11.7
Brown	Flat low-lying with deep, fertile soils overlying Quaternary deposits. Free draining.	4.6
Green	Flat, low-lying with deep, heavy clay soils. Poorly drained and often saturated.	25.4
Olive	Flat, low-lying coastal plains with fine silt to clay soils underlain by marine or estuarine sediments. Poorly drained.	8.4
Orange	Low-lying, thin, acid peaty soils overlying impermeable Eocene bedrock. Poorly drained and often saturated.	0.4
Pink	Moderately elevated with deep soils overlying impermeable bedrock. Mixed drainage, generally poor.	29.4
Purple	Lowland, moderately hilly, with peaty soils often overlying impermeable Palaeogene bedrock.	3.6
Red	Urban, variable soils and geology but minimum exposed soil surface and normally impermeable surfaces.	0.8
Yellow	Upland with very steep slopes, moorland peat overlying impermeable bedrock. High proportion of overland flow and near-surface lateral quickflow.	15.7

extent cattle grazing and are therefore used less intensively. This means that despite the high degree of slope, abundance of rainfall generating runoff averaging 1200–2000 mm per year, together with the relatively high proportion of overland flow and near-surface lateral quickflow as a function of thin soils and scarce vegetation that overlie impermeable bedrock, the high nutrient export potential of these landscapes is not translated into high nutrient flow rates. The geoclimatic regions coloured pink have a somewhat similar geographic distribution to the yellow regions, but occur more frequently (29.4% compared with 15.7%) and on moderately elevated (<150 m amsl) acidic soils underlain by impermeable metamorphic bedrock. Differences between both regions also include greater realised export rates for N and P due to the greater intensity of crop and grassland production and higher livestock densities commonly supported in these regions. The accumulation of N and P inputs here are likely to generate a substantial pool of N and P which has potential for export to adjacent waters in wet conditions.

The regions assigned as purple were classed as yellow in the original framework. However, in contrast to the revised yellow regions, these landscape units are typically hilly lowland (<300 m amsl) peat systems often but not always overlying impermeable Palaeogene rock that have developed in areas around estuaries, in valley-bottom fens and other topographic depressions. Poor drainage leads to the area becoming water-logged. Owing to their relative accessibility compared to upland peat areas, the purple regions are subject to increased agricultural intensity. However, relatively low nutrient export rates occur owing to nutrient retention within the developing peat layers and the tight cycling of any available nutrients within the system. Likewise and also reclassed from the original yellow units, the orange regions support high export potential, but low actual nutrient export rates as a result of peat soils overlying Eocene bedrock in the few remaining lowland moorland areas of the UK, with no fertiliser use, and extensive, low stocking density livestock grazing.

Low to moderate N and P flow potential is characteristic of the flat dry regions of the UK, classified here as brown, despite intensive arable production with associated high rates of fertiliser N and P applications to crops and grass. This reflects the fact that despite a high rate of N and P input to this landscape, the flatness of the topography, the low rates of runoff (<150 mm per year) and permeable soils generate a low actual rate of total nutrient export to waters. In comparison, N and P export is greater in the lowland hilly blue region, which is underlain by permeable

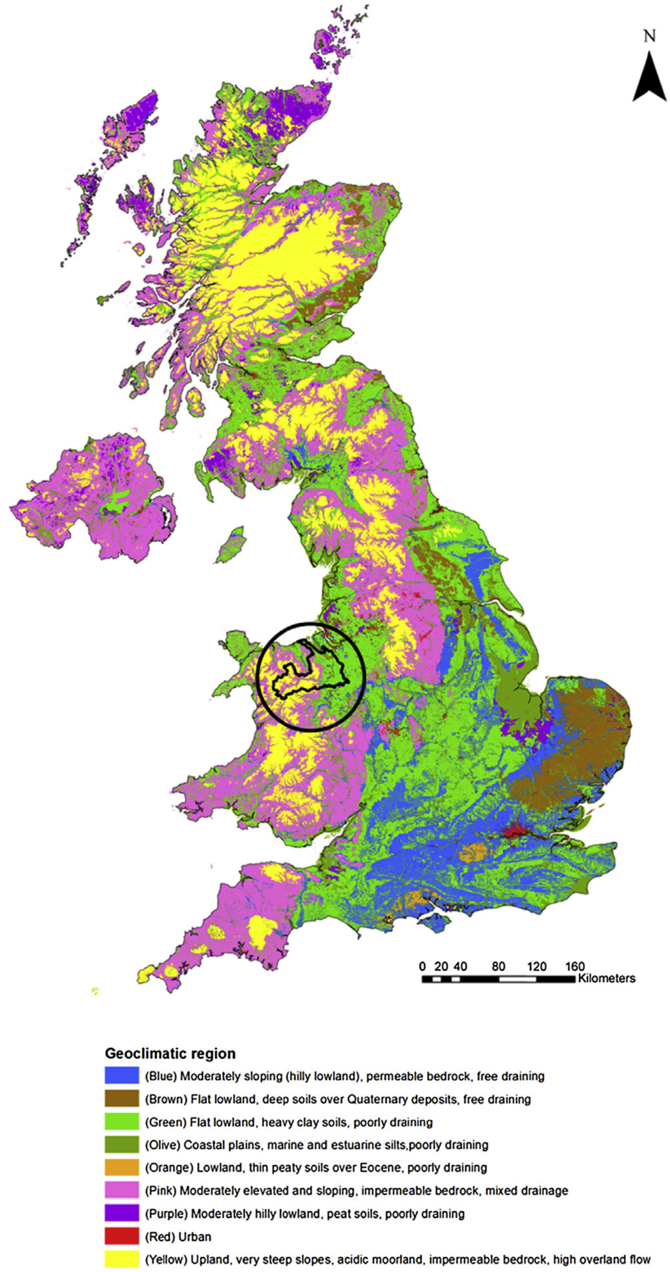


Fig. 3. The revised geoclimatic framework, extended to include Scotland and Northern Ireland and the first standardised nutrient modelling framework for the UK. The degree of spatial detail for a single catchment (River Dee, circled) is displayed in Fig. 5.

bedrock because of the high rate of N and P input, combined with the higher rate of nutrient export potential generated by steeper slopes, and higher rainfall intensity. This is reflected in the coefficient values assigned for this landscape type.

Finally, the highest rates of N and P export are predicted for the green and olive regions, representing flat lowland areas with heavy clay soils (green), river alluvium and flat coastal plains underlain by marine or estuarine sediments (olive). The olive regions are a subclass of the original green regions. These flat, wet, low permeability landscapes support the highest rates of grassland fertiliser application in the UK, together with the highest stocking densities for dairy and beef cattle production, which collectively lead to a combination of high N and P input rates and export potential. Much of this region is under-drained, facilitating rapid transport of nutrients and sediment to adjacent waters along the drainage system, and many of the soils are super-saturated with phosphorus following decades of intensive stock production and fertiliser application (Johnes and Butterfield, 2002; Johnes et al., 2007).

2.2.2. *Evaluation of the framework*

Bennett et al. (2013) recognise situations where a qualitative approach, rather than a quantitative method, is necessary for performance evaluation owing to the nature of the subject under review, such as participatory modelling frameworks for establishing a common understanding of a system. The geoclimatic framework was assessed using the best means available: a combined approach using a mixture of comparison with quantitative data (two forms of data independent to the framework training data) and a qualitative assessment using current knowledge of landscape character and land use.

First, the hydrological nature of the geoclimatic regions was investigated using the base flow index (BFI) and catchment descriptions of over 1400 hydrological catchments monitored in the national river flow archive (NRFA) at the Centre for Ecology and Hydrology (CEH). The BFI, ranging from 0 to 1.0 (Gustard et al., 1992) and derived using hydrograph analysis of long-term gauged daily mean flows, is a measure of the proportion of the river runoff that originates from stored sources. For example, rivers draining impervious clay catchments typically have low baseflow indices (0.15–0.35) and display an extreme flow regime. For these catchments, a high proportion of the flow in the river is delivered along overland and near-surface lateral quickflow pathways, particularly during storm events, affecting the timing, magnitude and chemistry of nutrient delivery to the water body (Johnes, 2007). Conversely, permeable chalk streams typically have a BFI greater than 0.9 as a result of a high proportion of the flow in the river delivered from groundwater sources (i.e. a large amount of runoff is routed through the slow response store). Permeable catchments convey greater dampening of the rainfall signal with larger within-catchment stores: moderate flows prevail with the absence of high peak flow events (Allen et al., 2014) and therefore also affect the transfer of nutrients to a waterbody (Evans et al., 2004; Evans and Johnes, 2004; Johnes, 2007; Yates and Johnes, 2013). For the purpose of evaluating the geoclimatic region framework, the BFI of catchments acted as an independent proxy for describing geology (Bloomfield et al., 2009) and the delivery of nutrient loads to waterbodies (Johnes, 2007; Deelstra et al., 2014). Catchment boundaries for the NRFA gauging points were overlaid on the revised geoclimatic regions and three catchments for each region that contained >85% composition of the catchment was selected for evaluation. A mean value for the assigned BFI for the three selected catchments for each geoclimatic region was compared for consistency with the inherent hydrological properties of the geoclimatic regions.

A second evaluation of the framework, in the form of grids, rather than catchments, was carried out using BFIHOST data. This base flow index is a measure of catchment hydrology derived using the 29-class hydrology of soil types (HOST) classification, a grouping of soil associations into classes based on physical properties of soils and on their hydrogeological setting (Gustard et al., 1992; Boorman et al., 1995). The BFIHOST values were derived from multivariate regression of the soil type data against long-term BFI data for representative catchments in the UK. The BFIHOST data comprises a 1 km^2 gridded spatial resolution and were overlaid on the geoclimatic region framework. The BFIHOST value that coincided with each grid cell in the geoclimatic region framework was extracted and the percentage composition in each region calculated.

The composition of geoclimatic regions present in a single river catchment, the River Dee that flows into England to the Irish Sea from its headwaters in Wales, was examined for consistency using qualitative data on catchment characteristics and land use (Natural Resources Wales, 2013).

2.3. *Model application*

The year 2000 was selected for demonstrating the function of the geoclimatic region framework with the supporting ECM. Digital data required for populating the model were acquired from a range of data sources, manipulated in ArcGIS Desktop 10.0 and transferred to a geospatial database for model application within the EVOp portal cyber-infrastructure.

2.3.1. *Data source*

Data for land use and livestock (total number of cattle, pigs, sheep and poultry) in Britain (England, Wales and Scotland) were taken from the Annual Agricultural Census Returns for 2000 at a spatial resolution of 4 km^2 grid. The data comprised the area of land used for cereal crops, other arable crops, bare fallow land, permanent grassland, temporary (ley or rotational) grassland, the area of rough grazing land (unfertilised), and the area of woodland/orchards. Corresponding data for Northern Ireland were sourced from the Northern Ireland Agricultural Census 2000. The total number of people in Britain and Northern Ireland were obtained from the respective 2001 Census of Human Population. Atmospheric N deposition across the UK data was sourced from the Department for Environment, Food and Rural Affairs (Defra) at a spatial resolution of 5 km^2. Owing to a deficiency of distributed P deposition data, a predefined value based on that reported by Johnes (1996) was adopted. This is an area of input data uncertainty to the model, requiring refinement through future research. Input coefficients to accompany the model data included values for the average amount of TN and TP produced per human/livestock unit (by type) annually, the nature and extent of sewage treatment facilities/livestock manure handling, fertiliser application rates to crops and grassland and N fixation by crop type (taken from the Survey of Fertiliser Practice), using and adapting coefficient ranges previously reported by Johnes et al. (2007) and Johnes and Butterfield (2002).

2.3.2. *Spatial database*

A spatial database was developed at a grid-based spatial scale of 4 km^2 across the whole of the UK, comprising 63 241 independent grid cells. Each 4 km^2 grid cell was given a unique identification number and the gridded structure was converted to an ESRI shapefile. Input parameters (e.g. permanent grass, urban area) were given a unique codename and parameter information was extracted per grid by intersection of the data shapefiles with the 4 km^2 grid shapefile. Following data population, a shapefile describing the model database and 4 km^2 grid boundaries was then overlaid on a shapefile containing the revised geoclimatic region framework. The proportion of each geoclimatic region in every 4 km^2 grid was extracted and a percentage composition of each region within the grid was assigned. The new shapefile (model database plus 4 km^2 grid boundaries) was then intersected with shapefiles representing the geographic boundaries of various water quality management reporting units (river catchment, Water Framework Directive River Basin District (WFD RBD), Coastal Drainage Unit, OSPAR marine zones, and UK countries). Columns were created in the database to indicate the spatial location of each 4 km^2 grid with respect to the various reporting units.

2.3.3. *Scenarios*

Distributed scenario testing of multiple potential land-based nutrient mitigation measures was carried out to explore their potential impact on the rate of nutrient export from land to water. The scenarios target key sources of nutrients in certain geoclimatic regions that threatened the potential for meeting water quality targets (Johnes et al., 2007). Measures to reduce P export to waters from point source discharges from wastewater treatment facilities, sewage works and septic tank systems were also included. The scenarios tested represent conditions appropriate under (1) Compliance with Good Agricultural Practice (GAP) policy guidance, (2) mitigation through on-farm measures, (3) mitigation measures appropriate to the delivery of reduced diffuse N and P loads to support Water Framework Directive (WFD) compliance. In addition to those tested under (1–3), (4) included measures to ensure compliance with the standards required under the European Union (EU) Urban Wastewaters Treatment Directive (UWwTD) to all sewage sources in the UK (80% reduction of P export from humans). Details on each scenario are presented under these headings in Table 2. Manipulations were performed on a geoclimatic region basis, using the concept of identifying problem sources for regions similar to Johnes et al. (2007).

Table 2
Description of scenario conditions.

#	Scenario	Condition
1	Good Agricultural Practice (GAP) policy guidance	Compliance with practices detailed in the UK Code of Good Agricultural Practice
2	Mitigation through on-farm measures	Reduce overall N and P export from sheep production by 50% + reduce N and P fertiliser application rates to all crops and grass by 50%
3	Farming for Water Framework Directive (WFD) compliance	Scenario 2 + reduce overall N and P export from cattle production systems by 25%
4	European Union (EU) Urban Wastewaters Treatment Directive (UWwTD) compliance	Scenario 3 + 80% reduction of P export from all point source sewerage systems through introduction of P stripping (N is not affected)

2.3.4. *Web application*

The ECM was deployed as a web application as part of the EVOp portal in order to gain wide accessibility, high availability, and good user experience. A web tool was developed (supported by the geoclimatic region framework) in order to provide the user with an interactive and intuitive interface that is accessible for any device with a modern web browser.[1]

The ECM implementation was imported from its previous format as a Microsoft Excel spreadsheet to a web application composed of two parts: a *cloud-hosted* part and an *in-browser* part. These are described below, after briefly introducing the underlying cloud infrastructure.

The EVOp cloud infrastructure was developed to enable hosting a myriad of models (e.g. hydrological, soil) and a large amount of data (including time series, geographical, topological, nutrients, etc.). This infrastructure is managed by a bespoke *Infrastructure Manager* that monitors computational load and usage and, accordingly, automatically enforces resource management rules to ensure high performance for all connected users and low operational cost (based on a predefined policy that is tailored to the specifics of the used cloud service providers). Modellers, in liaison with distributed system developers, use this infrastructure to construct web services that could be remotely triggered to process data. For the purposes of the ECM, the respective web service conforms to the Web Processing Services (WPS) standard (Castronova et al., 2013). This is an Open Geospatial Consortium (OGC) specification for how geospatial inputs and outputs should be handled between different computer software units.

The cloud-hosted part of the application handles the bulk of the calculations for the model for any general area of interest, enabling the application to leverage the computational processing power and disk space provided by the cloud. This part of the application is developed using MySQL for storing and querying spatial data, and PHP for exposing a WPS-compliant web service that is used for communication with the user's device. Due to the elastic computational capabilities offered by the cloud infrastructure, the service could be invoked using different input parameter values repeatedly by the same user or by a large number of different users at the same time.

The service passes summary tables back to the user's browser, at which point the in-browser code takes over the processing responsibilities. This code, written in JavaScript, runs in the user's web browser to perform some final calculations using the summary tables and visualise the output in an interactive manner that allows the user to zoom in and out of different parts of the results to explore different scenarios. A JavaScript framework (jQuery[2]) and FLOT[3] (a plotting library for jQuery) were also used for generating pie and barcharts. These were used in conjunction with HTML (HyperText Markup Language) and CSS (Cascading Style Sheets) which are used to structure and style the web pages. Performing the final stage of the calculation on the user's device facilitates scenario testing together with fast calculation and visualisation of the model estimates at any spatial scale. Calculations were performed for each 4 km^2 cell encompassed by the selected area, assuming that the land uses were assigned equally as per percentage composition of geoclimatic region in the cell. The user experience of the modelling process, including the method of scenario application, is demonstrated in the video http://vimeo.com/107474845. Fig. 4 is a screenshot of the output of the web tool that the user could use to further explore and manipulate the results. Further technical details are available in Elkhatib et al. (2013), Emmett et al. (2014) and Vitolo et al. (2015).

2.3.5. *Evaluation of model estimates*

Owing to a lack of comprehensive landscape scale data the modelled outcomes for nutrient loads per year were assessed using qualitative comparisons with other landscape scale model outcomes for the UK: (1) the integrated IDEA-INTEGRATOR model which has not been specifically calibrated to UK conditions but has been applied there at a 1 km^2 grid resolution (Leip et al., 2011) and (2) the Global News model (Seitzinger et al., 2010). In accordance with Bennett et al. (2013), the performance of modelling future conditions, such as the interventions applied as scenarios in this study, could not be directly assessed in quantitative terms as representative data do not exist. Nevertheless, lack of independent data for evaluation is not constraining as the purpose of the current research is to demonstrate the geoclimatic region framework, rather than to test the applicability of a nutrient model.

3. Results and discussion

3.1. *Geoclimatic framework evaluation*

The mean BFI of the three catchments comprising >85% composition of a single geoclimatic region class were consistent with the hydrological/permeability properties used for regional separation: 0.17 for green; 0.31 for yellow; 0.35 for pink; 0.62 for brown; and 0.89 for blue. There were no catchments for which the orange, olive or purple geoclimatic regions comprised >85% area. Evaluation also showed that BFIHOST classes coincided with the hydrological/permeability character of geoclimatic regions. For instance, 71% of HOST class 1 (BFI of 1.0) areas were comprised of the blue geoclimatic region, 60% of HOST class 19 (BFI of 0.4) areas contained the yellow geoclimatic region, and the green and olive geoclimatic made up 78% of the area contained by HOST class 9 (BFI of 0.35).

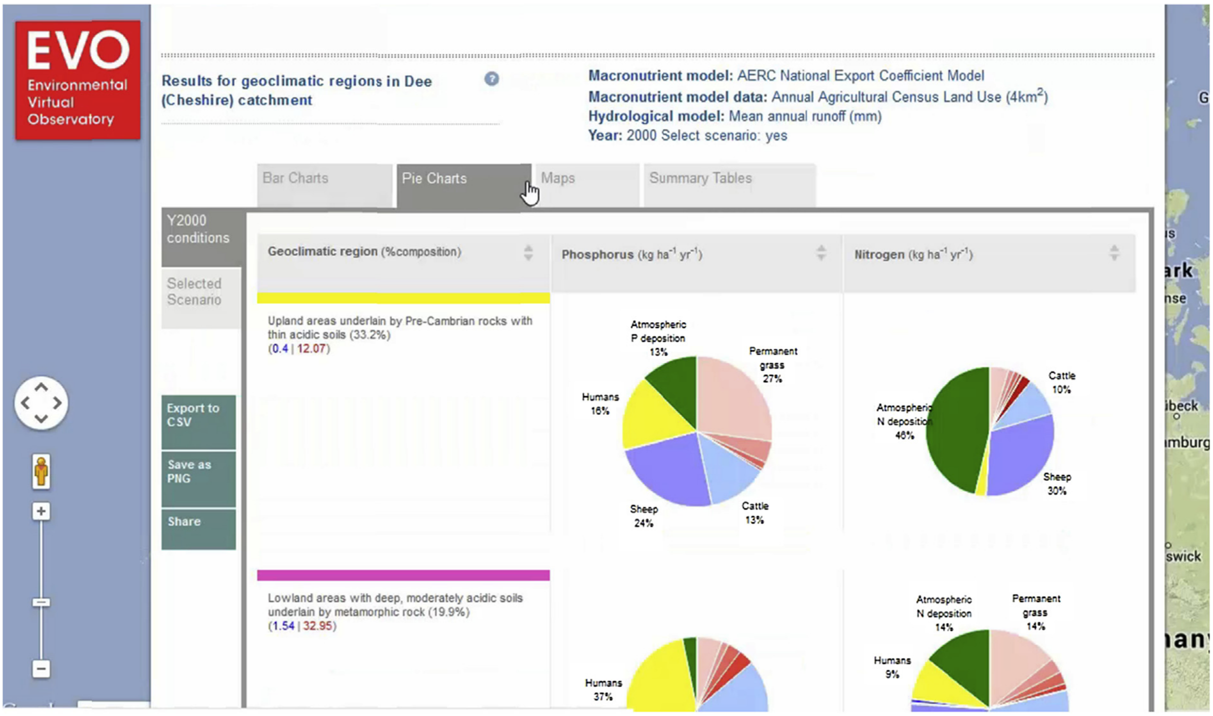
An example of the composition of the geoclimatic region framework for a single river catchment, the River Dee, is shown in

---

[1] This includes any of the major browsers released around early 2011, e.g. Chrome versions 9+, Firefox versions 6+, Safari versions 5+, Opera versions 10.50+.

[2] http://jquery.com, accessed 15th August 2014.

[3] http://www.flotcharts.org, accessed 15th August 2014.



Results for geoclimatic regions in Dee (Cheshire) catchment

Macronutrient model: AERC National Export Coefficient Model
Macronutrient model data: Annual Agricultural Census Land Use ($4km^2$)
Hydrological model: Mean annual runoff (mm)
Year: 2000 Select scenario: yes

Upland areas underlain by Pre-Cambrian rocks with thin acidic soils (33.2%)
(0.4 | 12.07)

Lowland areas with deep, moderately acidic soils underlain by metamorphic rock (19.9%)
(1.54 | 32.95)

Fig. 4. Screenshot of modelled output generated from running the geoclimatic region framework with the export coefficient model in the EVOp web tool. The modelled estimates are illustrated in pie-chart format to separate out various sources.

Fig. 5. The catchment of the River Dee is diverse both hydrologically and geologically; the landscape, therefore, varies considerably, and this is reflected in land use and human settlement patterns (Natural Resources Wales, 2013). The western upland part of the catchment on the edge of Snowdonia mountain range is predominately rural, with sheep farming on the poorer areas of grassland, with large areas of coniferous forestry plantation. This region is coded in the upland category (yellow). The mountains of this area reach altitudes that exceed 850 m and mean annual rainfall reaches 2500 mm. River flows rise quickly in response to precipitation events, rapidly returning to baseflow after the event (Outram et al., 2014). Moving east to the middle catchment, the landscape changes to lower elevation hill slopes supporting intensive mixed arable and livestock production (pink), while the lower reaches of the river drain through the Cheshire Plain, consisting of fertile clays (green) on which intensive dairy and beef cattle production is practised. The hydrological regime changes in these lowland areas, less rainfall occurs with mean annual rainfall dropping to 700 mm near the basin outlet. A damped flow regime with a moderate response to rainfall events then predominates. Internationally important mudflats (olive) are contained along the fringes of the Dee Estuary while only 6% of the catchment is urbanised, where the main urban centres (red) account for over 60% of the human population resident in the catchment.

3.2. Model estimates

The gridded output data for total loading of TN per hectare (TN ha^{-1} yr^{-1}) in the Dee catchment (Fig. 6) reflects the nutrient export potential conceptualised in the modelling framework; greater exports are shown for the green and pink regions reflecting high nutrient supply and export potential, while the yellow region displays much lower realised export.

Examples of the ECM output at various spatial scales in 2000 for total TN and for TP are presented in Figs. 7 and 8. Outputs are shown in order of increasing aggregated spatial area characterised by different hydrological accounting units, ranging from 4 km^2 grid scale up to the major reporting units for the UK.

The geoclimatic regions encapsulate the concept that land capability for nutrient uptake or export and the risk of adverse water quality vary geographically. Areas with common environmental characteristics (e.g. soils, geology, and microclimate) behave similarly, therefore facilitating a means for coordinated regionalised landscape-scale remediation to achieve water quality goals. The land management scenarios explored (Table 2) are simple exploratory exemplars, reflecting the general concept of reduction measures, but outputs highlight the real challenge for effective mitigation schemes to determine which strategies are most appropriate in each landscape type in the context of different stressors.

The model output for the upland yellow regions shows that these areas export the lowest load of nutrient per unit area, just 3% and 5% of total diffuse TP and TN loads of the total diffuse nutrient export estimated for the whole of the UK, (Tables 3 and 4) respectively, although the landscape type is the third most common in the UK (16%). These upland regions, often feeding into headwater streams, are nevertheless under pressure from diffuse pollution, and by acting to keep sediment and nutrients out of the stream's lower reaches plus supporting a range of species that are uniquely adapted to their habitat, small changes in nutrient concentrations may have significant impacts owing to their low resilience to environmental change (Reed et al., 2009). The largest loads of nutrients and most noticeable changes which result from the remediation scenarios occur in the areas largely dominated by the green and olive landscape type. The flat, impermeable lowland green regions make up 25% of the UK's land mass, although the export of nutrients from these areas provide a disproportionate 36% and 43% of the total TN and TP diffuse nutrient export estimated for the whole of the UK, respectively, highlighting the importance of spatial targeting with tailored interventions. The free-draining regions classed as blue and brown display moderate to high exports

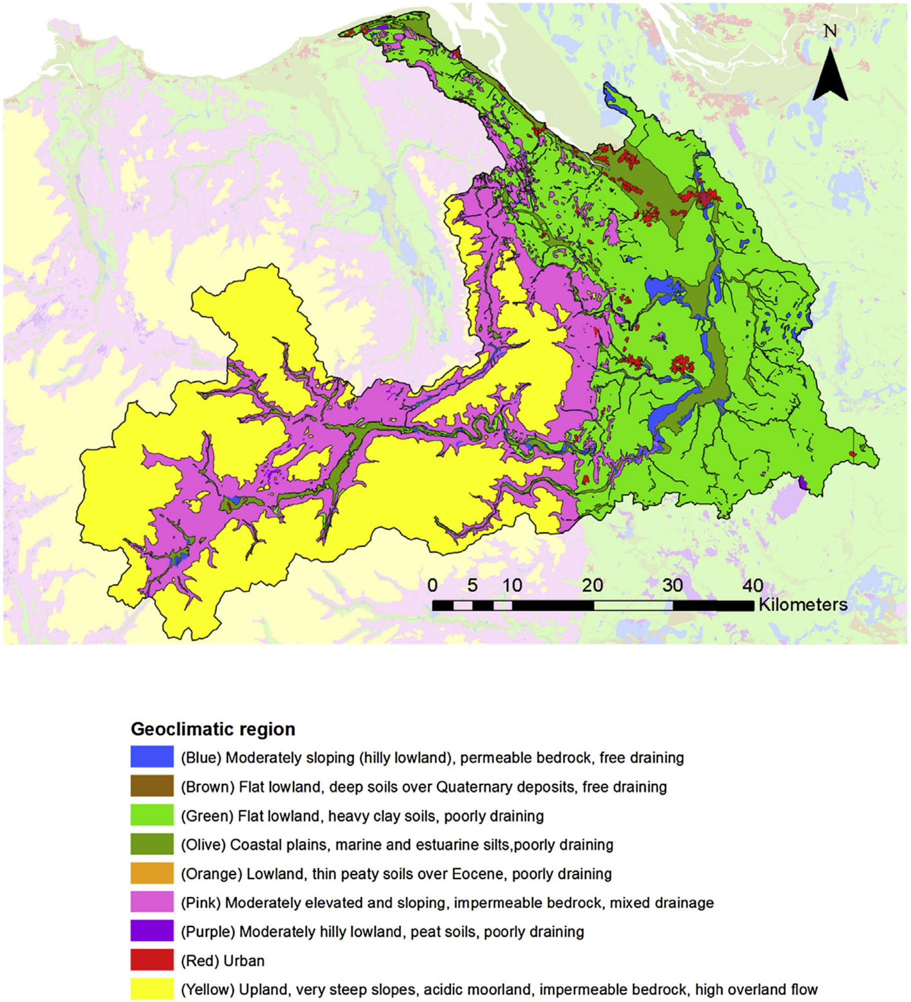


Fig. 5. Location and extent of the revised geoclimatic framework for the River Dee Water Framework Directive River Basin District (WFD RBD). The location of the River Dee catchment in the UK context is indicated in Fig. 3.

of TN and TP under current conditions. The nutrient loads in the blue regions not only pose a threat to the groundwater-fed chalk streams in the area, but also the economically important aquifers that underlie these areas as they are extremely vulnerable to contamination (Stuart et al., 2011; Wang et al., 2012, 2013). At present UWwTD compliance is only required for larger WwTW serving a population equivalent greater than or equal to 10 000 persons. The scenario testing here suggests that even with a selection of diffuse measures in place, a maximum of 58% reduction in P export can occur (Fig. 9), with the greatest rates of reduction in P export occurring in major urban centres. Primarily, the results show that targeting urban sources of TP is likely to demonstrate the most immediate decrease in nutrient loading. This contrasts markedly with the output from scenario testing for N (Fig. 10), with a maximum of only 29% reduction in N export even under the most stringent conditions tested, and the greatest rates of reduction in N export occurring in lowland agricultural areas, particularly in the impermeable green western areas where livestock farming on heavy clay soils is common.

3.2.1. Evaluation of model estimates

Calibrated model output from the ECM application shows annual TN flux rates ranging from <2 kg TN ha^{-1} yr^{-1} on the west coast of Scotland to >130 kg TN ha^{-1} yr^{-1} in the major urban centres in south-east England, with a national average rate of N flux in lowland rural catchments of approximately 40 kg TN ha^{-1} yr^{-1}, and lowest rates of total TN flux in East Anglia (25 kg N ha^{-1} yr^{-1}). The integrated IDEA-INTEGRATOR model, which has not been specifically calibrated to UK conditions but has been applied there at a 1 km^2 grid resolution (Leip et al., 2011), shows comparable ranges, with <2.5 kg ha^{-1} yr^{-1} in the Highlands and Islands of Scotland, > 100 kg ha^{-1} yr^{-1} in urban centres of south east England and upland areas,[4] with an average rate of N flux in rural lowland catchments of approximately 35 TN kg ha^{-1} yr^{-1}, and the lowest rates of N flux in East Anglia at approximately 20 TN ha^{-1} yr^{-1} (Leip et al., 2011). While there has been no attempt to explicitly compare the two models, the patterns and flux rates are broadly similar.

The Global News model attempts to predict TN flux at the river catchment scale in three forms: dissolved inorganic N (nitrate plus nitrite plus ammonium, DIN), dissolved organic N (DON), and particulate N. The results look impressive at a global scale, but when we break this down they offer a poor representation of flux patterns across the UK at the scale at which catchment

---

[4] There is more explicit treatment of N deposition in this approach than in the export coefficient modelling approach, hence the higher predicted rates of N flux in uplands receiving high rates of N deposition.

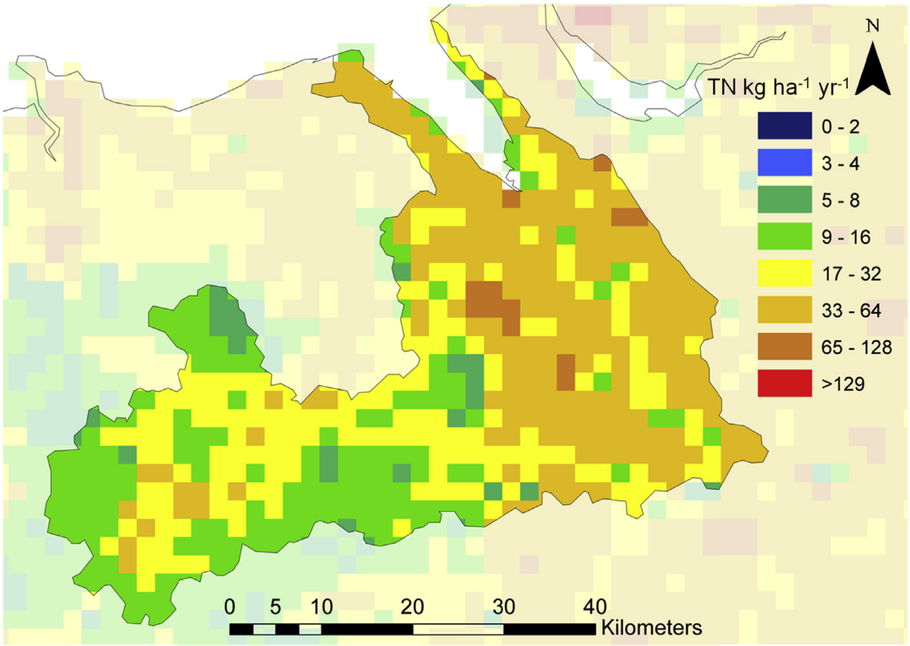


Fig. 6. Gridded output of total nitrogen (TN ha^{-1} yr^{-1}) export at 4 km^2 resolution for the year 2000 for the River Dee WFD RBD (Water Framework River Basin District) is shown. The location of the River Dee catchment in the UK context is indicated in Fig. 3.

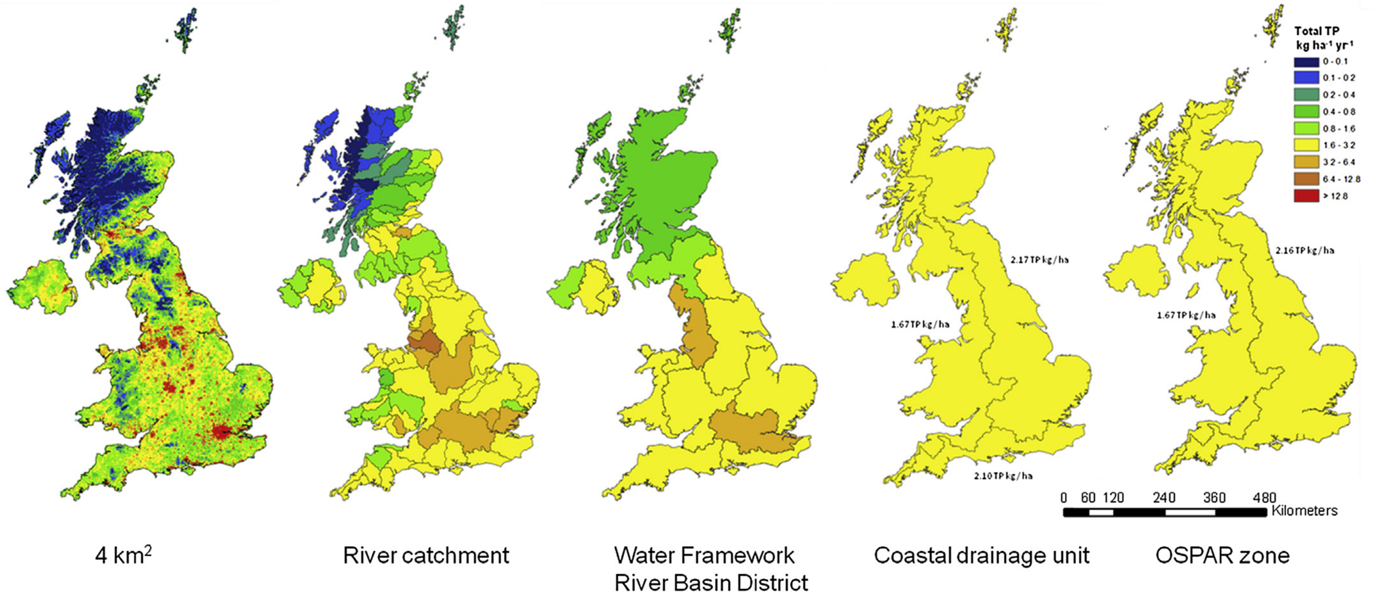
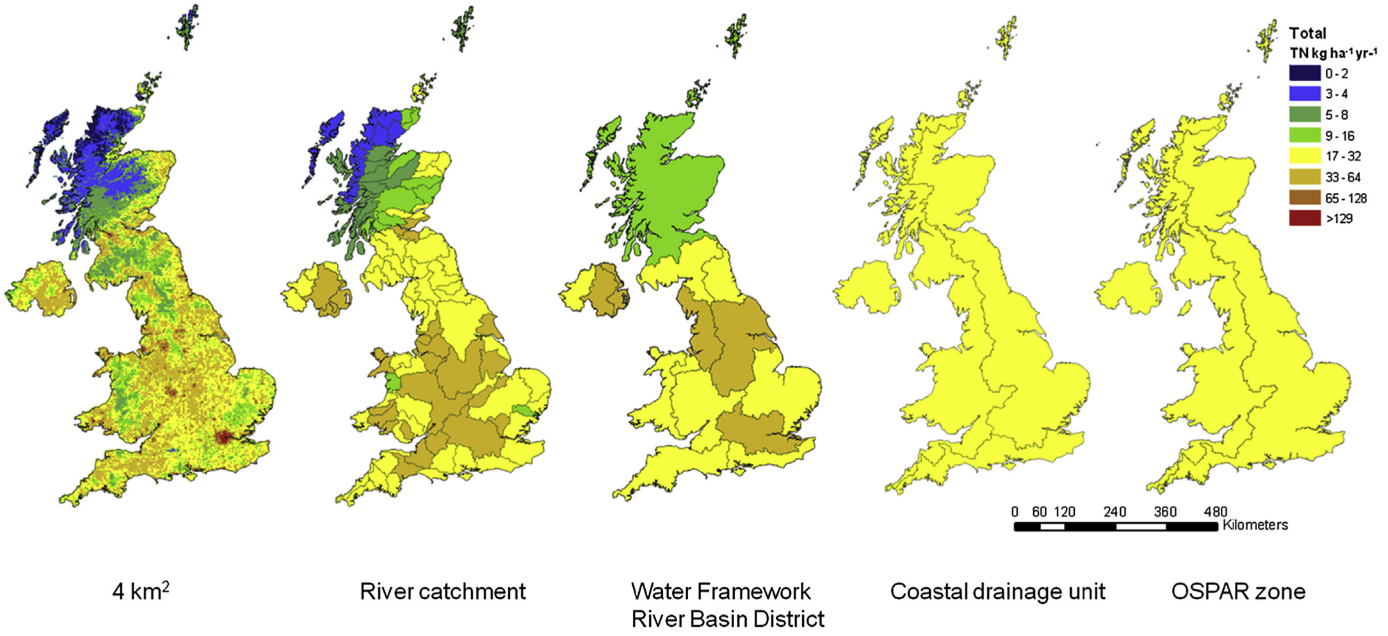


Fig. 7. Model outputs for total phosphorus (TP) flux (kg ha^{-1} yr^{-1}) across the UK using data for the year 2000. Outputs are shown in order of increasing spatial area, starting from 4 km^2 grids up to OSPAR marine zones. Data outputs for the coarser scale units were aggregated from the 4 km^2 grid scale outputs. Export in kg yr^{-1} for each grid square is aggregated across the region, with the final result reported in kg ha^{-1} yr^{-1}.

management takes place. Global News DIN flux for 2005 is estimated to lie in the range 1.9 (The Wash region) to 52.7 kg TN ha^{-1} yr^{-1}, with no discrimination of the west coast of Scotland from SE England. Even if the DON and particulate N fluxes were to be added to provide a total N flux estimate, both the range and variation in N flux estimates across the UK would provide a poor fit to observed N flux. This poor model fit to observed N flux behaviours derives primarily from the lack of a consideration for the variations in landscape within this model, and such model outcomes would provide unrealistic and misleading information to policy makers if disaggregated. The lack of a physical basis for this type of modelling provides apparent functionality, but lacks effective representation of environmental form and function and cannot be recommended as providing reliable underpinning for environmental policy and management at the catchment scale.

3.3. Implications of approach

Guidelines and recommendations for effective water resource management emphasise the importance of integrated approaches, cross-sectoral planning and participation of stakeholders and the public (European Commission, 2000; Rizzoli et al., 2008; Laniak



4 km^2 River catchment Water Framework River Basin District Coastal drainage unit OSPAR zone

Fig. 8. Model outputs for total nitrogen (TN) flux (kg ha^{-1} yr^{-1}) across the UK using data for the year 2000. Outputs are shown in order of increasing spatial area, starting from 4 km^2 grids up to OSPAR marine zones. Data outputs for each of the coarser units were aggregated from the modelled outputs at 4 km^2 scale. Export in kg yr^{-1} for each grid square is aggregated across the region, with the final result reported in kg ha^{-1} yr^{-1}.

et al., 2013). Although the focus in this paper has been on P and N in water, many of the methods described and discussion points relate to environmental data and models in general (e.g. carbon, ecological metrics, hydrology). In order to create a better understanding of the dominant processes and controls along environmental and spatial gradients it is important that novel and flexible modelling structures are developed and run efficiently. Approaches are also needed to provide scenarios testing capability to understand how environmental change and interventions may perturb cycles, and the likely efficacy of proposed mitigation measures. For instance, a low-entry barrier tool with the capability to undertake rapid exploratory analysis, such as the scenario analysis described here, is suitable for investigating the likely benefit of programmes of measures (POMs) for compliance with WFD requirements in Europe.

The need for integrated environmental modelling has been instigated by contemporary environmental problems and policies. The advancement of this approach is facilitated by transdisciplinary science and the strength of modern computers that allow the environment to be treated in a holistic way (Laniak et al., 2013). Integrated management for water resources is particularly suitable at the scale of the hydrological catchment; catchments typically present complex environments that comprise multiple processes and stakeholders. An increased emphasis on integrated water assessment and management at a catchment scale has led to the further development of modelling tools (Holguin-Gonzalez et al., 2013; Kelly et al., 2013; Lescot et al., 2013). For example hydrological science is transforming to meet the needs of sustainable water management strategies (Wagener et al., 2010). These

Table 3
Summary of model outputs for TN per geoclimatic region for the year 2000 and scenario runs.

Geoclimatic region	Current (2000) Total TN (kg ha^{-1})	Current (2000) Diffuse TN (kg ha^{-1})	% contribution to total diffuse load	Scenarios		
				1 Diffuse TN (kg ha^{-1})	2 Diffuse TN (kg ha^{-1})	3 Diffuse TN (kg ha^{-1})
Blue	46.22	18.98	10.88	18.61	14.19	13.27
(% difference from current)				(1.95)	(25.22)	(30.09)
Brown	25.28	16.23	3.67	15.92	14.60	13.82
(% difference from current)				(1.93)	(10.03)	(14.86)
Green	47.04	29.21	36.48	28.60	20.21	18.68
(% difference from current)				(2.09)	(30.83)	(36.06)
Olive	74.54	27.62	11.39	27.09	18.80	17.47
(% difference from current)				(1.92)	(31.93)	(36.74)
Orange	57.17	7.63	0.15	7.53	6.72	6.47
(% difference from current)				(1.31)	(11.84)	(15.11)
Pink	28.61	21.39	30.89	20.68	16.28	14.50
(% difference from current)				(3.32)	(23.93)	(32.22)
Purple	19.22	7.35	1.30	7.21	5.65	5.30
(% difference from current)				(1.92)	(23.14)	(27.94)
Yellow	8.05	6.56	5.07	6.49	5.53	5.37
(% difference from current)				(0.97)	(15.71)	(18.15)

Table 4
Summary of model outputs for total phosphorus (TP) per geoclimatic region for the year 2000 and scenario runs.

Geoclimatic region	Current (2000) Total TP (ha^{-1} yr^{-1})	Current (2000) Diffuse TP (ha^{-1} yr^{-1})	% Contribution to total diffuse load	Scenarios			
				1 Diffuse TP (ha^{-1} yr^{-1})	2 Diffuse TP (ha^{-1} yr^{-1})	3 Diffuse TP (ha^{-1} yr^{-1})	4 Total TP (ha^{-1} yr^{-1})
Blue	2.50	0.57	8.07	0.55	0.44	0.40	0.85
(% difference from current)				(2.88)	(22.74)	(29.99)	(65.89)
Brown	1.34	0.74	4.12	0.73	0.53	0.51	0.63
(% difference from current)				(1.08)	(27.88)	(30.55)	(52.93)
Green	2.81	1.41	43.48	1.35	1.16	1.03	1.31
(% difference from current)				(3.86)	(17.36)	(27.01)	(53.50)
Olive	2.57	1.36	13.93	1.32	1.03	0.91	1.51
(% difference from current)				(3.47)	(24.63)	(33.29)	(41.22)
Orange	3.29	0.18	0.09	0.18	0.15	0.14	0.76
(% difference from current)				(2.41)	(19.15)	(25.19)	(76.95)
Pink	1.39	0.73	26.15	0.70	0.57	0.49	0.62
(% difference from current)				(4.31)	(22.09)	(32.87)	(55.14)
Purple	0.41	0.26	1.13	0.25	0.19	0.18	0.21
(% difference from current)				(2.43)	(25.00)	(31.08)	(48.92)
Yellow	0.21	0.16	2.97	0.15	0.12	0.11	0.12
(% difference from current)				(1.83)	(23.94)	(28.52)	(42.37)

advancements present greater levels of integration between models and between modelling and decision making processes, leading to the development of better integrated environmental models (Beven, 2007) and the need for greater focus on evaluation (Matthews et al., 2011). The concept of the geoclimatic region framework presents an effective baseline layer on which biogeochemical model structures can be developed, and model estimates tested and compared in an impartial way. This concept of this tool can not only lead to a change in the way models are developed in the future but provides an effective way for current management to compare the strengths and weaknesses of models of various spatial resolution used to inform environmental policies at present. In the case of understanding and managing other environmental processes that have different controls and combination of pressures, the concept of the geoclimatic region framework can be applied with the manipulation of spatial data describing those factors.

The geoclimatic region framework, and its testing using a model structure is the first step in delivering a powerful tool for the prediction of nutrient loads under past, current and future conditions at a scale that can underpin national and international integrated nutrient management policy (e.g. WFD). However, the ECM is only one possible model suitable for this type of application in the geoclimatic modelling framework. Examples of other models suitable to adapt for the framework and to sit within the EVOp portal include SCIMAP (Reaney et al., 2011; Milledge et al., 2012), the INCA modelling suite (Whitehead et al., 1998; Wade et al., 2002), the INTEGRATOR (de Vries et al., 2003, 2011) and CAPRI DNDC European grid scale models (Leip et al., 2008, 2011), and INITIATOR (de

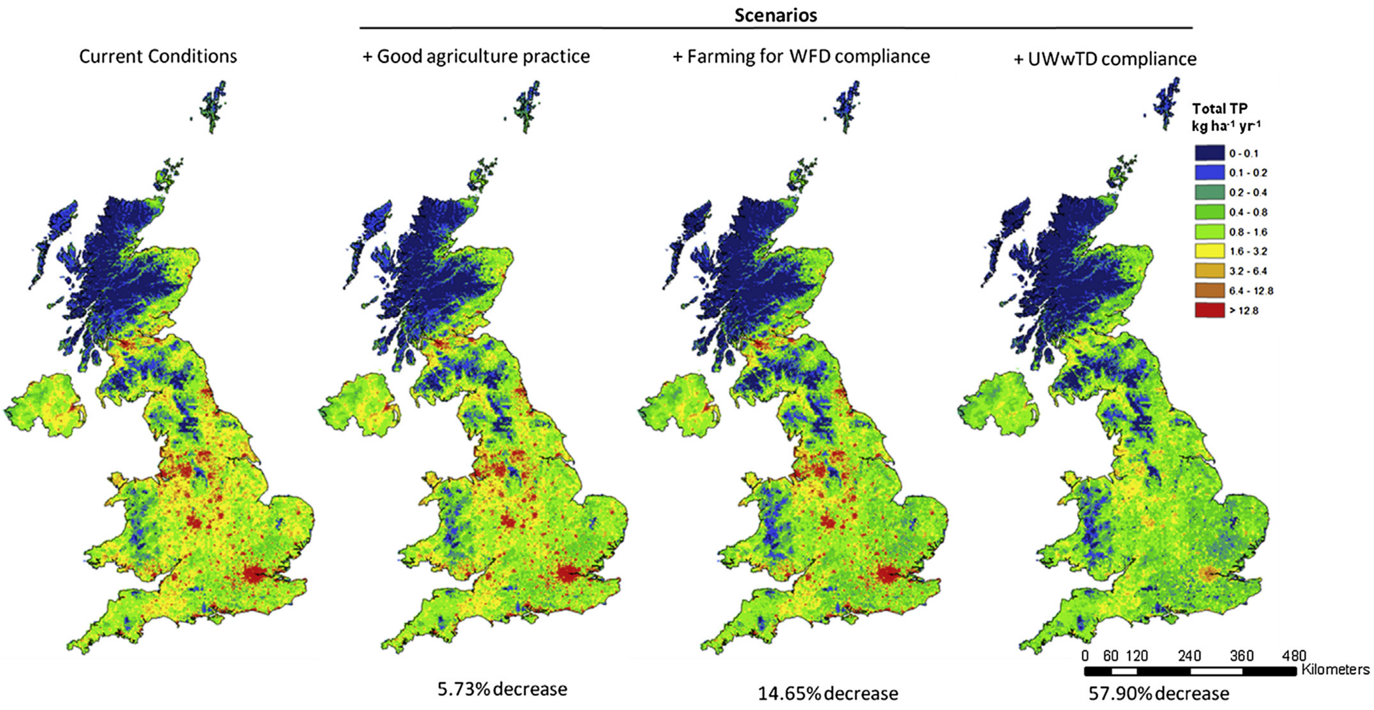


Fig. 9. Model outputs for total phosphorus (TP) flux (kg ha^{-1} yr^{-1}) based on nutrient management scenarios across the UK using data for the year 2000. The scenarios include reduction of point source TP generated from humans through the UWwTD compliance option.

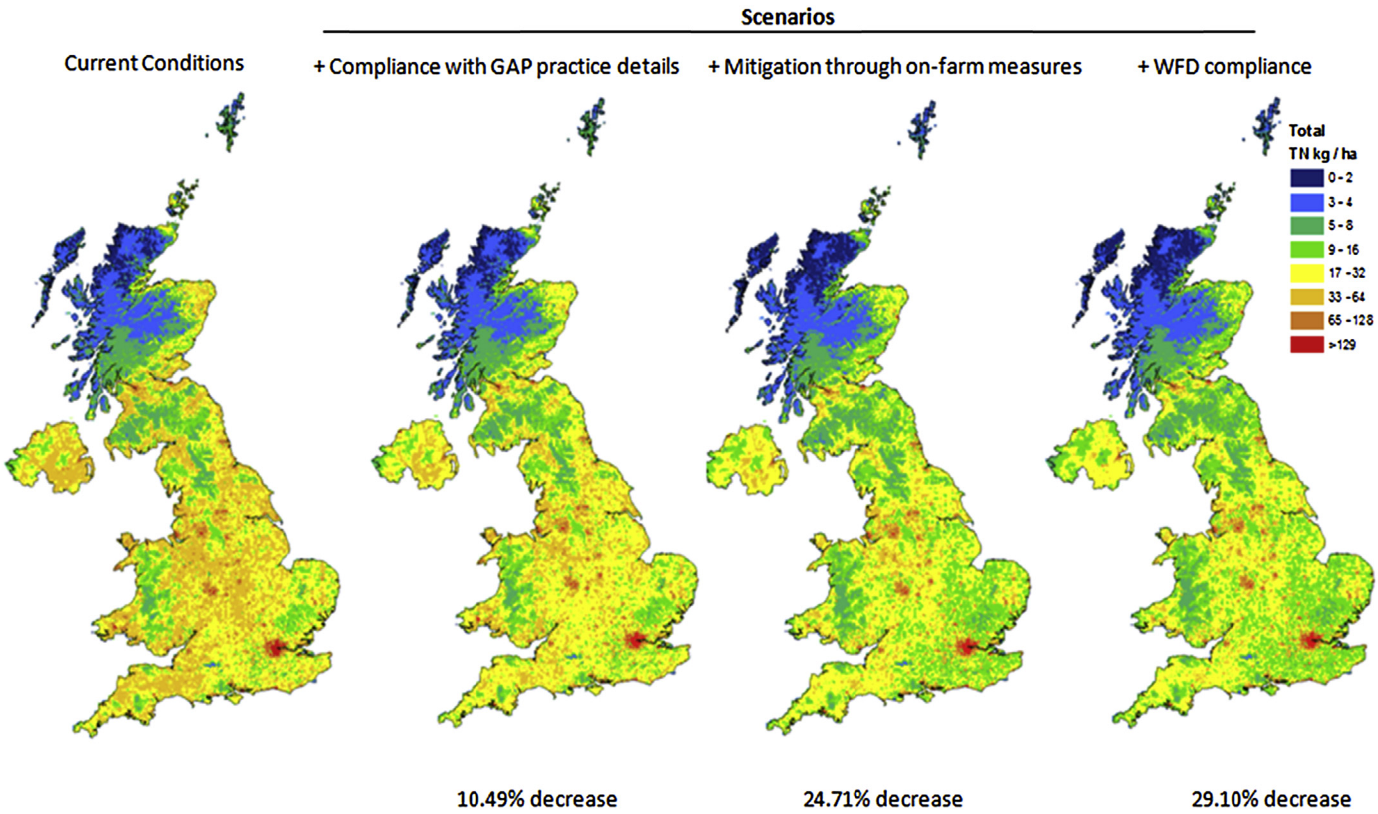


Fig. 10. Model outputs for total nitrogen (TN) flux (kg ha^{-1} yr^{-1}) based on nutrient management scenarios across the UK using data for the year 2000. The scenarios do not include reduction of point source TN generated from humans.

Vries et al., 2003). In order to fit into the framework models would require recalibration for the various geoclimatic regions. Models more complex than the ECM may be able to better account for explicit nutrient cycling and export behaviours over shorter time steps (e.g. seasonal or daily), intermediate processing of nutrients that may take place at different scales, and the rate of transfer of nutrients through a pathway of different geoclimatic regions to a waterbody. These aspects will be particularly important for assessing the importance or influence of policy-lead intervention scenarios. There is also the opportunity to adopt a group of independently developed models (i.e., as an ensemble modelling approach), such as the approach taken in a local hydrology component of the EVOp project (Emmett et al., 2014). The advantages of an ensemble modelling methodology include reducing the inherent uncertainties that individual model projections can produce by conveying the mean, range and probability distribution of the ensemble projections that would be useful to policy makers (Eilola et al., 2011; DeWeber and Wagner, 2014). Furthermore a multi-model approach may help identify and explore the spaces where the models have least agreement (Kronvang et al., 2009).

Global recognition of the benefits of a common platform for cross-disciplinary environmental data, models and decision support systems, together with the rapid evolution of geobrowsing technologies and familiarity with social networks, have driven key research initiatives such as the EVO in the UK and EarthCube, funded by the United States (US) National Science Foundation (NSF). By capitalising on the latest developments in cloud computing to efficiently enable a regionalised modelling framework, this work has provided the first test of the efficacy of running biogeochemical models with an elastic computational backend and with such a universal interface (as that provided by the cloud). The EVOp presents a low-entry barrier for a range of stakeholders with tools tailored to users with different levels of expertise and interest. The EVOp approach offers an effective tool for users to explore the likely impact of policy interventions and mitigation measures on nutrient loading to waters. Data can be shared, models tested and compared, scenario runs examined and results discussed, providing a robust, reproducible method for nutrient management and communication. The EVOp also presents a holistic learning environment, enabling transparent views of water quality monitoring and management to the whole community.

4. Conclusions

Sound spatial targeting of policy-based mitigation measures for preventing and remediating the impacts of pollution is essential for improving the quality of the environment and for increasing cost-effectiveness of such measures. However, environmental management at policy relevant spatial scales must be multifactorial, requiring knowledge on the myriad of sources and controls on the transfer of nutrients to receiving waterbodies. This study has shown that defining a restricted set of landscape elements reduces the dimensionality of such heterogeneity, thus minimising complexities and providing a reduced set of key features for consideration. The approach also facilitates the aggregation of data from grid plot to complex river catchments and, owing to the regionalisation of the structure, transfer of knowledge is efficiently conducted from data-rich areas to those lacking reliable observational data.

The geoclimatic region framework, run with an exemplar nutrient model in an exploratory cloud cyber-infrastructure, displays a powerful tool for the efficient estimation of nutrient loads at scales that can underpin national and international nutrient management policy. In terms of a practical and accessible management tool, the framework has many clear advantages over biogeochemical modelling approaches currently being used in water resource management. The time taken to run scenario analysis is also streamlined as a result of the framework structure and processing time of cloud computing. Development of integrated modelling capability through ensemble modelling may benefit as a result of the current research.

Acknowledgements

This work was supported by the United Kingdom Natural Environment Research Council (NERC) pilot projects on Environmental Virtual Observatory technologies NE/I002200/1 and NE/I004017/1 (Environmental Virtual Observatory pilot, EVOp). The authors acknowledge and would like to thank the members of the full project team (full list of team members in Emmett et al., 2014) for any contributions made to the research presented in this paper. Bloomfield and Lawley publish with the permission of the Executive Director of the British Geological Survey (NERC).

References

Allen, D.J., Darling, W.G., Davies, J., Newell, A.J., Gooddy, D.C., Collins, A.L., 2014. Groundwater conceptual models: implications for evaluating diffuse pollution mitigation measures. Q. J. Eng. Geol. Hydrogeol. 47 (1), 65–80. http://dx.doi.org/10.1144/qjegh2013-043.

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20. http://dx.doi.org/10.1016/j.envsoft.2012.09.011.

Bennion, H., Johnes, P., Ferrier, R., Phillips, G., Haworth, E., 2005. A comparison of diatom phosphorus transfer functions and export coefficient models as tools for reconstructing lake nutrient histories. Freshw. Biol. 50 (10), 1651–1670. http://dx.doi.org/10.1111/j.1365-2427.2005.01428.x.

Beven, K., 2007. Towards integrated environmental models of everywhere: uncertainty, data and modelling as a learning process. Hydrol. Earth Syst. Sci. 11 (1), 460–467. http://dx.doi.org/10.5194/hess-11-460-2007.

Beven, K.J., Alcock, R.E., 2012. Modelling everything everywhere: a new approach to decision-making for water management under uncertainty. Freshw. Biol. 57, 124–132. http://dx.doi.org/10.1111/j.1365-2427.2011.02592.x.

Bloomfield, J.P., Allen, D.J., Griffiths, K.J., 2009. Examining geological controls on baseflow index (BFI) using regression analysis: an illustration from the Thames Basin, UK. J. Hydrol. 373 (1–2), 164–176. http://dx.doi.org/10.1016/j.jhydrol.2009.04.025.

Boorman, D.B., Hollis, J.M., Lilly, A., 1995. Hydrology of Soil Types: a Hydrologically Based Classification of the Soils of the United Kingdom. Institute of Hydrology Report No. 126.

Burt, T.P., Johnes, P.J., 1997. Managing water quality in agricultural catchments. Trans. Inst. Br. Geogr. 21 (1), 61–68.

Castronova, A.M., Goodall, J.L., Elag, M.M., 2013. Models as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. Environ. Model. Softw. 41, 72–83. http://dx.doi.org/10.1016/j.envsoft.2012.11.010.

de Vries, W., Kros, J., Oenema, O., de Klein, J., 2003. Uncertainties in the fate of nitrogen II: a quantitative assessment of the uncertainties in major nitrogen fluxes in the Netherlands. Nutr. Cycl. Agroecosyst. 66 (1), 71–102. http://dx.doi.org/10.1023/A:1023354109910.

De Vries, W., Leip, A., Reinds, G.J., Kros, J., Lesschen, J.P., Bouwman, A.F., Grizzetti, B., Bouraoui, F., Butterbach Bahl, K., Bergamaschi, P., Winiwarter, W., 2011. Geographic variation in nitrogen budgets in agricultural and other terrestrial ecosystems over Europe. In: Sutton, M.A., et al. (Eds.), European Nitrogen Assessment. Cambridge University Press, Cambridge, UK, pp. 317–344.

Dean, S., Freer, J., Beven, K., Wade, A., Butterfield, D., 2009. Uncertainty assessment of a process-based integrated catchment model of phosphorus. Stoch. Environ. Res. Risk Assess. 23 (7), 991–1010. http://dx.doi.org/10.1007/s00477-008-0273-z.

Deelstra, J., Iital, A., Povilaitis, A., Kyllmar, K., Greipsland, I., Blicher-Mathiesen, G., Jansons, V., Koskiaho, J., Lagzdins, A., 2014. Hydrological pathways and nitrogen runoff in agricultural dominated catchments in Nordic and Baltic countries. Agric. Ecosyst. Environ. 195, 211–219. http://dx.doi.org/10.1016/j.agee.2014.06.007.

DeWeber, J.T., Wagner, T., 2014. A regional neural network ensemble for predicting mean daily river water temperature. J. Hydrol. 517, 187–200. http://dx.doi.org/10.1016/j.jhydrol.2014.05.035.

Eilola, K., Gustafsson, B.G., Kuznetsov, I., Meier, H.E.M., Neumann, T., Savchuk, O.P., 2011. Evaluation of biogeochemical cycles in an ensemble of three state-of-the-art numerical models of the Baltic Sea. J. Mar. Syst. 88 (2), 267–284. http://dx.doi.org/10.1016/j.jmarsys.2011.05.004.

Elkhatib, Y., Gordon, S.B., Bholanathsingh, S., 2013. Experiences of using a hybrid cloud to construct an environmental virtual observatory. In: Proceedings of the Third Workshop on Cloud Data and Platforms (a EuroSys 2013 Workshop), Prague, Czech Republic, April 14 2013.

Emmett, B.A., Gurney, R.J., McDonald, A.T., Blair, G., Buytaert, W., Freer, J., Haygarth, P., Johnes, P.J., Rees, G.H., Tetzlaff, D., Afgan, E., Ball, L.A., Beven, K., Bick, M., Bloomfield, J.B., Brewer, P., Delve, J., El-khatib, Y., Field, D., Gemmell, A.L., Greene, S., Huntingford, C., Mackay, E., Macklin, M.V., Macleod, K., Marshall, K., Odoni, N., Percy, B.J., Quinn, P.F., Reaney, S., Stutter, M., Surajbali, B., Thomas, N.R., Vitolo, C., Williams, B.L., Wilkinson, M., Zelazowski, P., 2014. NERC Environmental Virtual Observatory Pilot Final Report. Natural Environment Research Council, UK. NE/I002200/1.

European Commission, 2000. Directive 2000/60/EC of the European Parliament and of the Council establishing a framework for Community action in the field of water policy. Off. J. Eur. Commun. 43, 1–72.

Evans, D.J., Johnes, P.J., 2004. Physico-chemical controls on phosphorus cycling in two lowland streams. Part 1 – the water column. Sci. Total Environ. 329 (1), 145–163.

Evans, D.J., Johnes, P.J., Lawrence, D.S., 2004. Physico-chemical controls on phosphorus cycling in two lowland streams. Part 2–The sediment phase. Sci. Total Environ. 329 (1–3), 165–182. http://dx.doi.org/10.1016/j.scitotenv.2004.02.023.

Futter, M.N., Butterfield, D., Cosby, B.J., Dillon, P.J., Wade, A.J., Whitehead, P.G., 2007. Modeling the mechanisms that control in-stream dissolved organic carbon dynamics in upland and forested catchments. Water Resour. Res. 43 (2), W02424. http://dx.doi.org/10.1029/2006WR004960.

Greene, S., McElarney, Y.R., Taylor, D., 2013. A predictive geospatial approach for modelling phosphorus concentrations in rivers at the landscape scale. J. Hydrol. 504, 216–225. http://dx.doi.org/10.1016/j.jhydrol.2013.09.040.

Gustard, A., Bullock, A., Dixon, J.M., 1992. Low Flow Estimation in the United Kingdom. Institute of Hydrology Report 108. Wallingford, UK.

Haygarth, P.M., Page, T.J.C., Beven, K.J., Freer, J., Joynes, A., Butler, P., Wood, G.A., Owens, P.N., 2012. Scaling up the phosphorus signal from soil hillslopes to headwater catchments. Freshw. Biol. http://dx.doi.org/10.1111/j.1365-2427.2012.02748.x.

Holguin-Gonzalez, J.E., Everaert, G., Boets, P., Galvis, A., Goethals, P.L.M., 2013. Development and application of an integrated ecological modelling framework to analyze the impact of wastewater discharges on the ecological water quality of rivers. Environ. Model. Softw. 48, 27–36. http://dx.doi.org/10.1016/j.envsoft.2013.06.004.

Johnes, P., Moss, B., Phillips, G., 1996. The determination of total nitrogen and total phosphorus concentrations in freshwaters from land use, stock headage and population data: testing of a model for use in conservation and water quality management. Freshw. Biol. 36 (2), 451–473. http://dx.doi.org/10.1046/j.1365-2427.1996.00099.x.

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. J. Hydrol. 183 (3–4), 323–349. http://dx.doi.org/10.1016/0022-1694(95)02951-6.

Johnes, P.J., 1999. Understanding catchment history as a tool for integrated lake and catchment management. Hydrobiologia 395–396, 41–60.

Johnes, P.J., 2007. Meeting ecological restoration targets in European Waters: a challenge for animal agriculture. In: Swain, D., Charmley, E., Steel, J., Coffey, S. (Eds.), Redesigning Animal Agriculture: the Challenge for the 21st Century. CAB International, Wallingford, pp. 185–203.

Johnes, P.J., Butterfield, D., 2002. Landscape, regional and global estimates of nitrogen flux from land to sea: errors and uncertainties. Biogeochemistry 57–58 (1), 429–476. http://dx.doi.org/10.1023/a:1015721416839.

Johnes, P.J., Foy, R., Butterfield, D., Haygarth, P.M., 2007. Land use scenarios for England and Wales: evaluation of management options to support 'good ecological status' in surface freshwaters. Soil Use Manag. 23, 176–194. http://dx.doi.org/10.1111/j.1475-2743.2007.00120.x.

Kelly, R.A., Jakeman, A.J., Barreteau, O., Borsuk, M.E., ElSawah, S., Hamilton, S.H., Henriksen, H.J., Kuikka, S., Maier, H.R., Rizzoli, A.E., van Delden, H., Voinov, A.A., 2013. Selecting among five common modelling approaches for integrated environmental assessment and management. Environ. Model. Softw. 47, 159–181. http://dx.doi.org/10.1016/j.envsoft.2013.05.005.

Kronvang, B., Behrendt, H., Andersen, H.E., Arheimer, B., Barr, A., Borgvang, S.A., Bouraoui, F., Granlund, K., Grizzetti, B., Groenendijk, P., Schwaiger, E., Hejzlar, J., Hoffmann, L., Johnsson, H., Panagopoulos, Y., Lo Porto, A., Reisser, H., Schoumans, O., Anthony, S., Silgram, M., Venohr, M., Larsen, S.E., 2009. Ensemble modelling of nutrient loads and nutrient load partitioning in 17 European catchments. J. Environ. Monit. 11 (3), 572–583.

Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A., 2013. Integrated environmental modeling: a vision and roadmap for the future. Environ. Model. Softw. 39, 3–23. http://dx.doi.org/10.1016/j.envsoft.2012.09.006.

Lawley, R., 2011. The Soil-parent Material Database: a User Guide. British Geological Survey Open Report, OR/08/034. 53.

Lazar, A.N., Butterfield, D., Futter, M.N., Rankinen, K., Thouvenot-Korppoo, M., Jarritt, N., Lawrence, D.S.L., Wade, A.J., Whitehead, P.G., 2010. An assessment of the fine sediment dynamics in an upland river system: INCA-Sed modifications and implications for fisheries. Sci. Total Environ. 408 (12), 2555–2566. http://dx.doi.org/10.1016/j.scitotenv.2010.02.030.

Leip, A., Achermann, B., Billen, G., Bleeker, A., Bouwman, A., de Vries, W., Dragosits, U., Doring, U., Fernall, D., Geupel, M., Herolstab, J., Johnes, P., Le

Gall, A.C., Monni, S., Neveceral, R., Orlandini, L., Prud'homme, M., Reuter, H., Simpson, D., Seufert, G., Spranger, T., Sutton, M., van Aardenne, J., Voss, M., Winiwarter, W., 2011. Integrating nitrogen fluxes at the European scale. In: Sutton, M., et al. (Eds.), The European Nitrogen Assessment. Cambridge University Press, Cambridge, pp. 345–376.

Leip, A., Marchi, G., Koeble, R., Kempen, M., Britz, W., Li, C., 2008. Linking an economic model for European agriculture with a mechanistic model to estimate nitrogen and carbon losses from arable soils in Europe. Biogeosciences 5 (1), 73–94. http://dx.doi.org/10.5194/bg-5-73-2008.

Lescot, J.-M., Bordenave, P., Petit, K., Leccia, O., 2013. A spatially-distributed cost-effectiveness analysis framework for controlling water pollution. Environ. Model. Softw. 41, 107–122. http://dx.doi.org/10.1016/j.envsoft.2012.10.008.

Liu, C., Kroeze, C., Hoekstra, A.Y., Gerbens-Leenes, W., 2012. Past and future trends in grey water footprints of anthropogenic nitrogen and phosphorus inputs to major world rivers. Ecol. Indic. 18, 42–49. http://dx.doi.org/10.1016/j.ecolind.2011.10.005.

Matthews, K.B., Rivington, M., Blackstock, K., McCrum, G., Buchan, K., Miller, D.G., 2011. Raising the bar? – the challenges of evaluating the outcomes of environmental modelling and software. Environ. Model. Softw. 26 (3), 247–257. http://dx.doi.org/10.1016/j.envsoft.2010.03.031.

McGonigle, D.F., Harris, R.C., McCamphill, C., Kirk, S., Dils, R., Macdonald, J., Bailey, S., 2012. Towards a more strategic approach to research to support catchment-based policy approaches to mitigate agricultural water pollution: a UK case-study. Environ. Sci. Policy 24, 4–14. http://dx.doi.org/10.1016/j.envsci.2012.07.016.

Milledge, D.G., Lane, S.N., Heathwaite, A.L., Reaney, S.M., 2012. A Monte Carlo approach to the inverse problem of diffuse pollution risk in agricultural catchments. Sci. Total Environ. 433, 434–449. http://dx.doi.org/10.1016/j.scitotenv.2012.06.047.

Natural Resources Wales, 2013. Dee River Basin Management Plan: SEA Scoping Document, June 2013.

Outram, F.N., Lloyd, C.E.M., Jonczyk, J., Benskin, C.M.H., Grant, F., Perks, M.T., Deasy, C., Burke, S.P., Collins, A.L., Freer, J., Haygarth, P.M., Hiscock, K.M., Johnes, P.J., Lovett, A.L., 2014. High-frequency monitoring of nitrogen and phosphorus response in three rural catchments to the end of the 2011–2012 drought in England. Hydrol. Earth Syst. Sci. 18 (9), 3429–3448. http://dx.doi.org/10.5194/hess-18-3429-2014.

Pärn, J., Pinay, G., Mander, Ü., 2012. Indicators of nutrients transport from agricultural catchments under temperate climate: a review. Ecol. Indic. 22, 4–15. http://dx.doi.org/10.1016/j.ecolind.2011.10.002.

Reaney, S.M., Lane, S.N., Heathwaite, A.L., Dugdale, L.J., 2011. Risk-based modelling of diffuse land use impacts from rural landscapes upon salmonid fry abundance. Ecol. Model. 222 (4), 1016–1029.

Reed, M.S., Bonn, A., Slee, W., Beharry-Borg, N., Birch, J., Brown, I., Burt, T.P., Chapman, D., Chapman, P.J., Clay, G.D., Cornell, S.J., Fraser, E.D.G., Glass, J.H., Holden, J., Hodgson, J.A., Hubacek, K., Irvine, B., Jin, N., Kirkby, M.J., Kunin, W.E., Moore, O., Moseley, D., Prell, C., Price, M.F., Quinn, C.H., Redpath, S., Reid, C., Stagl, S., Stringer, L.C., Termansen, M., Thorp, S., Towers, W., Worrall, F., 2009. The future of the uplands. Land Use Policy 26 (Suppl. 1), S204–S216. http://dx.doi.org/10.1016/j.landusepol.2009.09.013.

Rizzoli, A.E., Leavesley, G., Ascough Ii, J.C., Argent, R.M., Athanasiadis, I.N., Brilhante, V., Claeys, F.H.A., David, O., Donatelli, M., Gijsbers, P., Havlik, D., Kassahun, A., Krause, P., Quinn, N.W.T., Scholten, H., Sojda, R.S., Villa, F., 2008. Integrated modelling frameworks for environmental assessment and decision support (Chapter Seven). In: Jakeman, A.J., Voinov, A.A., Rizzoli, A.E., Chen, S.H. (Eds.), Developments in Integrated Environmental Assessment, vol. 3. Elsevier, pp. 101–118.

Robson, B.J., 2014. State of the art in modelling of phosphorus in aquatic systems: review, criticisms and commentary. Environ. Model. Softw. 61, 339–359. http://dx.doi.org/10.1016/j.envsoft.2014.01.012.

Seitzinger, S.P., Mayorga, E., Bouwman, A.F., Kroeze, C., Beusen, A.H.W., Billen, G., Van Drecht, G., Dumont, E., Fekete, B.M., Garnier, J., Harrison, J.A., 2010. Global river nutrient export: a scenario analysis of past and future trends. Glob. Biogeochem. Cycles 24 (2). http://dx.doi.org/10.1029/2009GB003587 pdf.

Stuart, M.E., Gooddy, D.C., Bloomfield, J.P., Williams, A.T., 2011. A review of the impact of climate change on future nitrate concentrations in groundwater of the UK. Sci. Total Environ. 409 (15), 2859–2873. http://dx.doi.org/10.1016/j.scitotenv.2011.04.016.

Sutton, M.A., Oenema, O., Erisman, J.W., Leip, A., Van Grinsven, H., Winiwarter, W., 2011. Too much of a good thing. Nature 472 (7342), 159–161.

Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J.A., Buytaert, W., 2015. Web technologies for environmental Big Data. Environ. Model. Softw. 63 (0), 185–198. http://dx.doi.org/10.1016/j.envsoft.2014.10.007.

Wade, A., Whitehead, P., Butterfield, D., 2002. The Integrated Catchments model of Phosphorus dynamics (INCA-P), a new approach for multiple source assessment in heterogeneous river systems: model structure and equations. Hydrol. Earth Syst. Sci. 6 (3), 583–606.

Wagener, T., Sivapalan, M., Troch, P.A., McGlynn, B.L., Harman, C.J., Gupta, H.V., Kumar, P., Rao, P.S.C., Basu, N.B., Wilson, J.S., 2010. The future of hydrology: an evolving science for a changing world. Water Resour. Res. 46 (5), W05301. http://dx.doi.org/10.1029/2009WR008906.

Wang, L., Butcher, A.S., Stuart, M.E., Gooddy, D.C., Bloomfield, J.P., 2013. The nitrate time bomb: a numerical way to investigate nitrate storage and lag time in the unsaturated zone. Environ. Geochem. Health 35 (5), 667–681. http://dx.doi.org/10.1007/s10653-013-9550-y.

Wang, L., Stuart, M.E., Bloomfield, J.P., Butcher, A.S., Gooddy, D.C., McKenzie, A.A., Lewis, M.A., Williams, A.T., 2012. Prediction of the arrival of peak nitrate concentrations at the water table at the regional scale in Great Britain. Hydrol. Process. 26 (2), 226–239. http://dx.doi.org/10.1002/hyp.8164.

Whitehead, P.G., Wilson, E.J., Butterfield, D., 1998. A semi-distributed Integrated Nitrogen model for multiple source assessment in Catchments (INCA): part I — model structure and process equations. Sci. Total Environ. 210–211 (0), 547–558. http://dx.doi.org/10.1016/S0048-9697(98)00037–0.

Yates, C.A., Johnes, P.J., 2013. Nitrogen speciation and phosphorus fractionation dynamics in a lowland Chalk catchment. Sci. Total Environ. 444 (0), 466–479. http://dx.doi.org/10.1016/j.scitotenv.2012.12.002.

Ye, S., Covino, T.P., Sivapalan, M., Basu, N.B., Li, H.Y., Wang, S.W., 2012. Dissolved nutrient retention dynamics in river networks: a modeling investigation of transient flows and scale effects. Water Resour. Res. 48 http://dx.doi.org/10.1029/2011wr010508.