National scale evaluation of the InVEST nutrient retention model in the United Kingdom

National scale evaluation of the InVEST nutrient retention model in the United Kingdom

John W. Redhead [ad]
Linda May [b]
Tom H. Oliver [a1]
Perrine Hamel [c]
Richard Sharp [c]
James M. Bullock [a]

[a] NERC Centre for Ecology and Hydrology, Maclean Building, Wallingford, Oxfordshire, OX10 8BB, UK.
[b] NERC Centre for Ecology and Hydrology, Bush Estate, Penicuik, Midlothian, EH26 0QB, UK
[c] Natural Capital Project, Woods Institute for the Environment, Stanford University, 371 Serra Mall, Stanford, CA 94305, USA
[d] Corresponding author: Email: johdhe@ceh.ac.uk, Tel: (+44) 1491 692538, Fax: (+44) 1491 692424
[1] Present address: School of Biological Sciences, Harborne Building, University of Reading, Reading, Berkshire RG6 6AS, UK

Abstract

A wide variety of tools aim to support decision making by modelling, mapping and quantifying ecosystem services. If decisions are to be properly informed, the accuracy and potential limitations of these tools must be well understood. However, dedicated studies evaluating ecosystem service models against empirical data are rare, especially over large areas. In this paper, we report on the national-scale assessment of a new ecosystem service model for nutrient delivery and retention, the InVEST Nutrient Delivery Ratio model. For 36 river catchments across the UK, we modelled total catchment export of phosphorus (P) and/or nitrogen (N) and compared model outputs to measurements derived from empirical water chemistry data.

The model performed well in terms of relative magnitude of nutrient export among catchments (best Spearman's rank correlation for N and P, respectively: 0.81 and 0.88). However, there was wide variation among catchments in the accuracy of the model, and absolute values of nutrient exports frequently showed high percentage differences between modelled and empirically-derived exports (best median absolute percentage difference for N and P, respectively: ± 64%, ± 44%). The model also showed a high degree of sensitivity to nutrient loads and hydrologic routing input parameters and these sensitivities varied among catchments.

These results suggest that the InVEST model can provide valuable information on nutrient fluxes to decision makers, especially in terms of relative differences among catchments. However, caution is needed if using the absolute modelled values for decision-making. Our study also suggests particular attention should be paid to researching input nutrient loadings and retentions, and the selection of appropriate input data resolutions and threshold flow accumulation values. Our results also highlight how availability of empirical data can improve model calibration and performance assessment and reinforce the need to include such data in ecosystem service modelling studies.

Keywords

Ecosystem services, nutrient delivery, runoff, eutrophication, river, land cover

Abbreviations

BRE - Beale ratio estimator, CEH-GEAR – centre for ecology and hydrology gridded estimates of areal rainfall, DEM - digital elevation model, IHDTM - integrated hydrological digital terrain model, InVEST - integrated valuation of ecosystem services and tradeoffs, LCM2007 - land cover map 2007, LULC - Land use/land cover, NDR - nutrient delivery ratio, NRFA - national river flow archive, TFA - threshold flow accumulation, UKCP09 - UK climate projections, WIMS - water information management system, WWTW - wastewater treatment works

1. Introduction

The ecosystem services concept is increasingly widely applied by decision makers seeking to assess the likely impacts of environmental change on human health and wellbeing (Braat and de Groot 2012; Tallis et al. 2008). For ecosystem services to be useful in practice, they must be quantified and mapped to identify the risks, impacts and potential trade-offs associated with predicted or known environmental change, or among different change scenarios (Malinga et al. 2015). To achieve such assessments, a wide variety of methods and tools have been developed to map, quantify and value the provision of ecosystem services (Fisher et al. 2009; Malinga et al. 2015; Seppelt et al. 2011; Sharps et al. 2017).

In recent years, many ecosystem service modelling tools have become freely available to the global user community. This overcomes issues surrounding proprietary software and data formats, and enables model development and application to benefit from increased data and model sharing, cloud computing facilities and a larger user community (Feng et al. 2011). Critically, these tools model multiple services, allowing users to take a multi-criterion approach to decision-making (Keller et al. 2015). Whilst the free and open-source nature of such tools brings many advantages, it allows users to run a wide range of models, and obtain results, with little knowledge of the modelling process or expertise in the subject area. A potential pitfall is that users may not familiarise themselves with the intended use and limitations of the model before using it, and may be unaware of the uncertainty associated with results that they incorporate into decision making processes (Willcock et al. 2016). Whilst a body of literature has begun to emerge exploring the strengths and weaknesses of these models (Dennedy-Frank et al. 2016; Redhead et al. 2016; Sharps et al. 2017; Willcock et al. 2016) the number of studies seeking to validate and explore the sensitivities of ecosystem service models remains limited (Hamel et al. 2017; Maes et al. 2012; Malinga et al. 2015; Schulp et al. 2014; Seppelt et al. 2011), especially over the large (i.e. regional to national) spatial scales at which much resource management policy is formulated (e.g. Wilby et al. 2006). Such studies are vital in providing user communities with the information required to choose the tools that are most appropriate for their particular situation, to use them correctly, and to understand associated uncertainties (Willcock et al. 2016). They can also provide valuable information on potential data sources for parameterising models, and help to focus data acquisition by revealing which parameters have the most influence on model accuracy. As a result, recent reviews have identified that one of the key obstacles to successful ecosystem service mapping and implementation into decision making processes is the comparative scarcity of validation or measurements of uncertainty in many applications of ecosystem service models (Maes et al. 2012; Malinga et al. 2015; Schulp et al. 2014; Seppelt et al. 2011)

Freshwater ecosystem service models that assess how land management affects water quantity and quality have the advantage of using physical variables that are commonly used in hydrologic modelling, even though these contribute to a wide range of different final services, from recreation to human health (Keeler et al. 2012). One of the most frequently modelled services is nutrient retention, which represents the reduction in nutrient loads between sources and receiving watercourses, due to biogeochemical processes involved in nutrient transport. Models of nutrient retention (e.g InVEST, ARIES, LUCI (Sharps et al. 2017; Vigerstol and Aukema 2011)), typically use a hydrologic module representing nutrient retention processes or, where available, direct outputs from more complex nutrient models (e.g. SWAT, RHESSys, see reviews in Breuer et al. (2008); and Shepherd et al. (1999)). When the modelling approach includes quantitative estimates of nutrient transport and retention, it becomes comparatively easier to validate models, because measurements of water chemistry are, in many countries, collected by environmental bodies and the water industry and these can be used to estimate watercourse loads for comparison with model outputs. Whilst this approach falls short of measuring a final ecosystem service (Keeler et al. 2012), it is an important step in providing the biophysical underpinning for any further assessments of ecosystem service value.

In this study, we used data from UK national monitoring to perform a thorough evaluation of the recently released nutrient retention tool of the Integrated Valuation of Ecosystem Services and Tradeoffs (InVEST, Sharp et al. 2016) ecosystem service modelling suite. InVEST is widely used for modelling multiple ecosystem services and considering trade-offs (e.g. Bai et al. 2013; Leh et al. 2013; Nelson et al. 2009; Sánchez-Canales et al. 2012; Sharps et al. 2017) and is free and open-source. We used national scale, spatially distributed data (of the sort available to most potential users) for model inputs and performed validation against a long-term, empirically-measured dataset. Our objectives were 1) to examine the sensitivity of the model to variation in input parameter values, spatial resolution and data sources, and 2) to determine the accuracy of the model against empirical data when using the most informative combination of input parameter values, for both phosphorus (P) and nitrogen (N).

2. Methods

2.1. The InVEST nutrient delivery ratio model

The InVEST (v.3.3.3) suite of tools has been developed to enable decision makers to assess trade-offs across ecosystem services and to compare the consequences of different future change scenarios, for example in land use or climate (Sharp et al. 2016). To this end, InVEST comprises a set of models that cover a wide range of ecosystem services. Like many ecosystem service models, these models are based on comparatively simple production functions, enabling them to be run quickly on a standard desktop computer and to take advantage of readily available data (Sharp et al. 2016) and targeting a user community with potentially limited technical background.

The UK has a long history of issues arising from nutrient contamination of watercourses (Johnes et al. 1996; Withers and Lord 2002), as it is densely populated and has a large proportion of its land area under anthropogenic land uses (i.e. agricultural and urban land). This results in high levels of nutrient input to freshwater systems, and ensuing concerns over the contamination of drinking water and damage to aquatic ecosystems via eutrophication (Withers and Lord 2002). Validated nutrient export models, with clear estimates of their accuracy and uncertainty are therefore particularly valuable to compare nutrient exports under different scenarios of environmental change or management interventions over larger spatial scales (Johnes et al. 1996; Shepherd et al. 1999; Wilby et al. 2006).

The InVEST nutrient delivery ratio (NDR) model aims to quantify relative nutrient export and retention across different catchments or sub-catchments, and to reflect changes in nutrient export/retention under different change scenarios. The model maps the transport of nutrients from catchment sources to the stream network. It combines the advantages of nutrient transport models (e.g. SWAT (Arnold et al. 1998); RHESSys (Tague and Band 2004)), which often work at the scale of subwatersheds or hydrological units to provide quantitative estimates of nutrient flows, and index models (Drewry et al. 2011), which spatially map source risk and transport factors.

The model computes a nutrient mass balance that represents the long-term, steady-state flow of nutrients based on i) nutrient sources associated with different land use/land cover (LULC) in the landscape, and ii) the retention properties (e.g. LULC, slope) of pixels belonging to the same flow path (Parn et al. 2012; Sharp et al. 2016). Specifically, nutrient sources across the landscape are derived from LULC-specific nutrient application (loading) rates, which can be determined from empirical data. Nutrient sources can be divided into surface and subsurface sources (which conceptually represent sediment-bound and dissolved components, a distinction common to many nutrient transport models (Newham et al., 2004; Newham et al., 2008). The model only includes diffuse sources of nutrient; point sources are not included and need to be added in post-processing of model outputs. Next, the model uses topographic routing and an index, the NDR factor, to emulate the movement of nutrients across the landscape and into a watercourse. The NDR factor is calculated for each landscape pixel based on the properties (e.g. slope, retention coefficient) of pixels that belong to the same flow path. This empirical approach is in contrast to more complex, process based models that incorporate detailed representations of nutrient cycling (see Breuer et al. 2008 for a review). At the catchment outlet, the nutrient export to water is calculated as the sum of the pixel-level contributions. For further details on the model, see Supplementary Material, Appendix S1 and Sharp et al. (2016). Model source code is available in Hamel and Sharp (2017)

Because of the qualitative nature of the NDR factor approach, calibration of the model is necessary to gain confidence in the quantitative outputs. The main calibration factor is the k_b parameter, which governs the relationship between the connectivity index, which is a function of topography, and the NDR factor. This relationship is further described in the user's guide (Sharp et al., 2016) and is akin to the structure of the InVEST sediment delivery ratio model (Hamel et al. 2015), which can be used independently to model this other facet of water quality.

2.2. Model Inputs

Spatially explicit model inputs required for the NDR model are a digital elevation model (DEM), land use/land cover (LULC) raster data, nutrient runoff proxy raster data and a vector delineation of the watersheds. We used the Centre for Ecology & Hydrology's Integrated Hydrological Digital Terrain Model (CEH IHDTM, Morris and Flavin 1990) for the DEM. The IHDTM was resampled or aggregated to the required resolution (see below), filled to eliminate sinks and combined with a digital watercourse network (Moore et al. 1994) to ensure routing along known watercourses. These processes were performed in ArcMap (v10.3 © ESRI, Redlands, CA). The model also requires a threshold value for flow accumulation (TFA) to define streams, which is expressed as a number of upstream pixels. Within the model, watercourses are assumed not to retain or add to the nutrient load, and nutrients reaching a stream pixel will contribute directly to the total load from the catchment (Sharp et al. 2016). The TFA value was selected following sensitivity analyses and examination of watercourse maps (See below, section 2.3).

LULC data were obtained from the 25 m resolution raster version of the UK Land Cover Map 2007 (LCM2007, Morton et al. 2011). The LCM2007 data are derived from satellite imagery, generalised digital cartography and image segmentation, and classify the UK land surface into 23 broad habitat classes (Jackson 2000; Morton et al. 2011). The InVEST model requires several parameter values for each distinct LULC class. These include the nutrient load applied to the land (kg ha^{-1} y^{-1}), the proportional retention of that nutrient load, the length of flow path required to achieve that retention (in metres), and the proportion of the nutrient load that travels via subsurface flow. This last variable is set to zero by default, making the assumption that all nutrients travel via surface or shallow subsurface flow. However, if modified, the model then requires two further parameters – the subsurface nutrient retention efficiency and the flow length required to achieve this.

Nutrient loading and nutrient retention coefficients for each LULC class were obtained by performing an extensive literature search for values relevant to the UK and for habitats that most closely matched the broad habitats defined by the LCM2007 (Supplementary Material, Table S1). Where several possible values for a single LULC class were found, the median value was used. A wide variety of sources provided information on P (Dillon and Kirchner 1975; Fozzard et al. 1999; Johnes 1996; May et al. 2001; May et al. 1996; McGuckin et al. 1999; Smith et al. 2005) with rather fewer supplying suitable values for N (Johnes 1996; Shi et al. 2006). Because many of these publications report measured or estimated export coefficients from land to water, which are a function of the two required model inputs (load to land and retention), some loads were estimated from export coefficients according to the following formula (Sharp et al. 2016):

$$Load\ to\ land = \frac{Export\ from\ land}{1 - Retention}$$

Critical flow length (i.e. the distance of travel required to achieve the nutrient retention coefficient) was set to the resolution of the input LULC raster across all LULC classes, catchments and nutrients, which was consistent with the relatively coarse resolution (25m at the minimum).

Previous studies have shown that choice of input data can have major impacts on the accuracy of InVEST ecosystem service models where these data relate to parameters to which the model is highly sensitive (Hamel and Guswa 2015; Pessacg et al. 2015; Redhead et al. 2016; Sánchez-Canales et al. 2012). We compared three sets of input data for the nutrient runoff proxy raster. These were, 1) WorldClim precipitation data (Hijmans et al. 2005), which are readily available, widely used and have global coverage interpolated to approximately 1km resolution 2) UK Met Office UKCP09 data at 5km resolution (Jenkins et al. 2008; Perry and Hollis 2005), which gave good estimates of total annual water yield when used in the relevant InVEST model (Redhead et al. 2016), and 3) CEH-GEAR data at 1km resolution (Tanguy et al. 2014), which has a higher spatial resolution. All datasets comprise gridded rainfall per raster cell at monthly or annual time steps, derived from interpolation and correction for geographic and topographic factors of measurements taken from a national network of meteorological stations. Data were derived from the mean of annual values between 2000 and 2012 to match the period of the validation data. We also tested a randomised dataset using values drawn from the range of all three datasets to test the impact of large errors in the nutrient runoff proxy raster on model accuracy.

2.3. Sensitivity Analysis

As well as varying input datasets for the nutrient runoff proxy raster we also tested the sensitivity of the model to changes in the values of the input parameters. This is key to understanding why the model behaves as it does, setting appropriate ranges for calibration of parameter values and helping subsequent users to identifying those parameters for which it is most worthwhile investing in to obtain more accurate data. To do this, first we ran the model on “hypothetical” versions of our test catchments, with the UKCP09 precipitation data, default values for threshold flow accumulation and k_b parameter (TFA = 1000 and k_b= 2, respectively), input LULC and DEM raster resolution of 25m and a single land cover class with a mean nutrient load to land (4.7 kg $ha^{-1}y^{-1}$) and retention (0.3) (because the model has the same structure, these analyses are valid for N and P). We then varied each of the precipitation data, nutrient load and nutrient retention by ±50% and ±90% and examined the percentage difference in modelled nutrient export to water. These values were chosen because the percentage difference between the median and maximum/minimum export coefficients was approximately 100%, so these variations explore the likely range of variation encountered when using literature derived coefficients.

For the single-value parameters (TFA and k_b) we explored a range of values. We tested three TFA values (100, 1,000 and 10,000). We used these three values because preliminary analyses determined that more subtle variations in TFA made very little difference to the overall length of steam network, especially in larger catchments. Preliminary analyses also determined that values below 100 were very likely to overestimate the stream network density, whilst values above 10,000 were not met in all catchments (i.e. no modelled watercourses were created). Because the ideal TFA value was catchment specific (see Results, section 3.1), we also used another approach, which involved setting the threshold either at default (1000) or high (10,000) but combining known watercourses into the LULC raster as a separate class with appropriately low retention. We used the same digital watercourse network to do this as was used to correct the flow paths generated from the DEM (Moore et al. 1994). For k_b we compared values of 0.5, 1, 2 (the default), 4, 8 and 16. Preliminary analyses determined that, whilst k_b is dimensionless and can in principle accept any value, values above this range made progressively less differences to the relationship between topography and nutrient delivery, whilst values below this range tend to collapse the function to the point where extreme changes in connectivity are required to impact on nutrient delivery. In all model runs we assumed a subsurface flow proportion of zero (i.e. all nutrient transported via surface flow).

Because the spatial scale and resolution of the input data can affect ecosystem service model outputs (Sharp et al. 2016), especially those with a dynamic flow component (Grafius et al. 2016), we also compared models run with versions of the LCM2007 and IHDTM at the highest resolution available (25m, the resolution of the raster LCM2007), and at lower resolutions that could easily be derived from these data (50m, the resolution of the IHDTM, 100, 200, 400 and 800 m.) Coarser resolutions greatly speed up the modelling but potentially reduce accuracy. When changing the resolution of the input rasters, TFA was adjusted to keep the flow path length consistent across raster resolutions, following Hamel et al. (2017). Coarser inputs than 800m were not tested, as at values above this some smaller catchments begin to have flow paths of only 1 or 2 cells, making setting an appropriate TFA impossible.

2.4. Validation data

The data used for validation were derived from the UK Environment Agency's Water Information Management System (WIMS), which provides records of total N and total P concentrations for a network of sampling points across England and Wales (Envrionment Agency 2017). Because these data represent instantaneous concentrations of nutrients, it was necessary to find sites with coincident records of river flow, and sufficiently frequent measurements of nutrient concentrations to enable the robust estimation of total annual nutrient load in the watercourse – comparable with the output of the NDR model – and to account for inter- and intra- annual variation. To achieve this, sites from WIMS were filtered to exclude sites with less than 5 years of available data over the years 2000-2010, with each year containing at least one measurement per month of total N or P. These sites were then overlain with the locations of all flow gauging stations in the National River Flow Archive (NRFA). The NRFA collates, quality controls, archives and disseminates hydrometric data from gauging stations operated by government funded environmental bodies across the UK (Fry and Swain 2010). WIMS sites that were spatially coincident with NRFA gauging stations had the necessary daily flow data available to enable annual nutrient loads to be calculated and their catchments had been previously defined using the IHDTM. These temporal and spatial filters resulted in 33 catchments being identified as having sufficient data to act as a validation dataset for P. However, because total N was measured at a smaller proportion of sites (most measure NOx), only three catchments met all of the above criteria for N. Therefore, we reduced to three the required number of years with at least monthly measurements, giving 16 catchments with sufficient data for N.

Total annual nutrient load for each year was calculated from the WIMS and NRFA data for each catchment using the Beale Ratio Estimator (BRE, Beale 1962) which relates the ratio of average load to average flow, at times when concentrations are measured, to the ratio of average true load to average true flow over the entire period of interest (Dunn et al. 2014). Whilst there are a wide variety of methods available with which to extrapolate loads from intermittent data, ratio estimators have been used in previous validation studies (Terrado et al. 2014) and the BRE has been shown to produce robust results, especially when the measurement frequency of the concentration data is lower than that for discharge (Dolan et al. 1981; Dunn et al. 2014; Meals et al. 2013; Quilbé et al.

2006; Richards and Holloway 1987), as was the case here. The median BRE nutrient load across years for each catchment was then calculated.

Because the NDR model only accounts for nutrients from diffuse sources, it was necessary to adjust the modelled output of total load by an estimated load for point sources, to enable comparison with the validation data. In the UK, point sources can contribute the majority of P and a substantial proportion of N to waterways (Edwards and Withers 2008), although this varies across space and time (Arheimer and Lidén 2000). The estimated load from point sources was obtained using a GIS layer of wastewater treatment works (WWTWs) provided from UK Water Companies through the Environment Agency (see Williams et al. 2009). Although there is a wide variety of other point sources of N and P releases (Edwards and Withers 2008), WWTWs are likely to be the largest contributor at a whole-catchment scale in the UK (Bowes et al. 2005; Edwards and Withers 2008). For each WWTW, data were available describing the maximum human population served and the treatment type employed (i.e. primary, secondary or tertiary). These data were combined with a mean annual per capita export of P and N in untreated sewage of 0.52 kg P and 4.5 kg N and nutrient retention efficiencies for the different treatment types, both derived from a recent UK-wide review (Naden et al. 2016), to give an estimated annual N and P output for each WWTW. N and P outputs from individual WWTWs were then summed to give an annual load from WWTWs per catchment. This value was then subtracted from the per-catchment BRE to give a total export from diffuse sources only for comparison with the output of the InVEST NDR model. We removed catchments for which the estimated nutrient export from point sources contributed to more than 50% of the total estimated export (mostly relatively heavily urbanised catchments, Fig. 1), as these were unlikely to be well represented by the model (which focuses on diffuse sources) and would be highly influenced by any errors in our estimation of point source nutrient exports, giving final sample sizes of 28 for P and 14 for N (Figure 1 and Supplementary Material, Table S1).

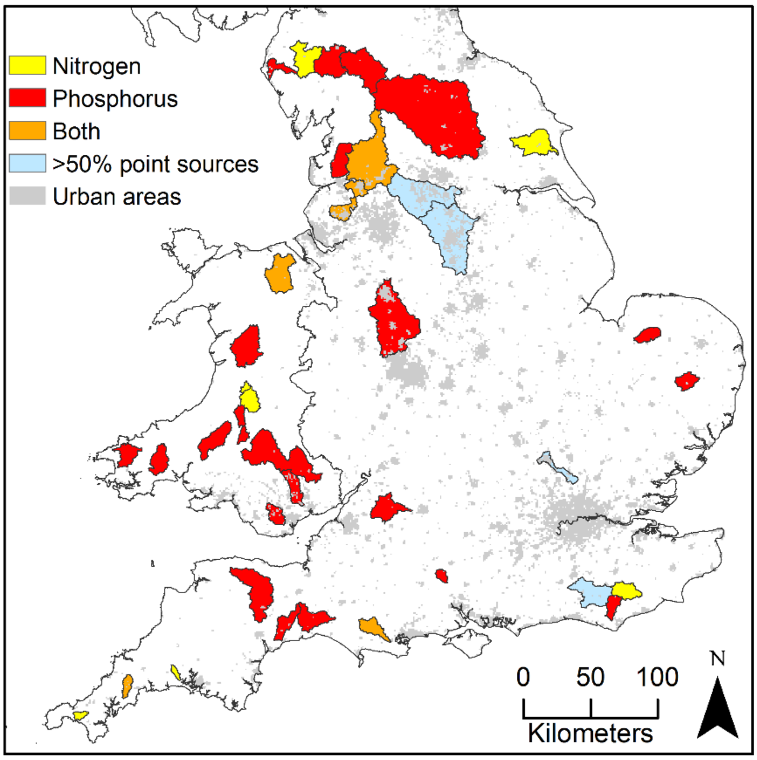

Fig. 1. Map of southern UK showing catchments providing validation data for nitrogen (yellow), phosphorus (red) or both (orange). Blue catchments indicate those which had sufficient nutrient and flow measurements, but were estimated to have over 50% of total nutrient runoff due to point sources and so were excluded from further analyses. Urban areas are also shown in grey (from LCM2007). Note that none of these catchments overlap.

2.5. Statistical Analysis

Comparisons between the modelled and measured data were made by performing linear regressions implemented in R (R Core Team 2014), as well as comparing the percentage differences between modelled and measured. Many stakeholders require models simply to predict accurately the rank order of locations in terms of ecosystem services, rather than absolute values (Willcock et al. 2016) and the InVEST model does not necessarily aim for accurate prediction of values (Sharp et al. 2016). Therefore, we also tested the accuracy of the InVEST NDR model in predicting relative export values using rank correlation (Spearman's rho).

3. Results

3.1. Sensitivity Analysis

Modelled nutrient export from the NDR model was insensitive to variation in precipitation (Supplementary Material Fig S1A). This was expected since these variations were applied as consistent percentage change across the entire spatial extent. Because the role of this input is to represent relative runoff between pixels, the model is still likely to be sensitive to different inputs where they show different spatial patterns, as opposed to different magnitudes. This was addressed by comparing the three different input datasets (see below, Section 3.2).

The model was sensitive to variation in the nutrient loading and retention values (Supplementary Material Fig S1B and S1C) although sensitivity was linear. Because land cover was held constant for these analyses, sensitivity to these parameters did not show any catchment specificity.

In contrast, sensitivity to the two calibration parameters was highly catchment specific. Figure 3 illustrates the percentage change in modelled nutrient export compared to the values obtained when using the default parameter values of 2 for k_b and 1000 for TFA. The effect of k_b on the magnitude and direction of change in nutrient export was catchment specific (Fig 2A). Overall, decreasing k_b to 0.5 produced the most extreme changes (-20% to +35%), whilst increasing k_b to 4 resulted in changes of ±10%. Further increases in k_b resulted in changes that remained within this range for the majority of catchments (Fig 2A). Catchment sensitivity appeared driven by topography, with more topographically varied catchments in the uplands showing decreases in nutrient export in response to increased k_b values and less varied, lowland catchments showing the opposite response (Pearson's r against % change at k_b = 0.5; Mean catchment altitude n = 35, r = 0.704, $p < 0.001$; Standard deviation in catchment altitude n = 35, r = 0.709, $p < 0.001$).

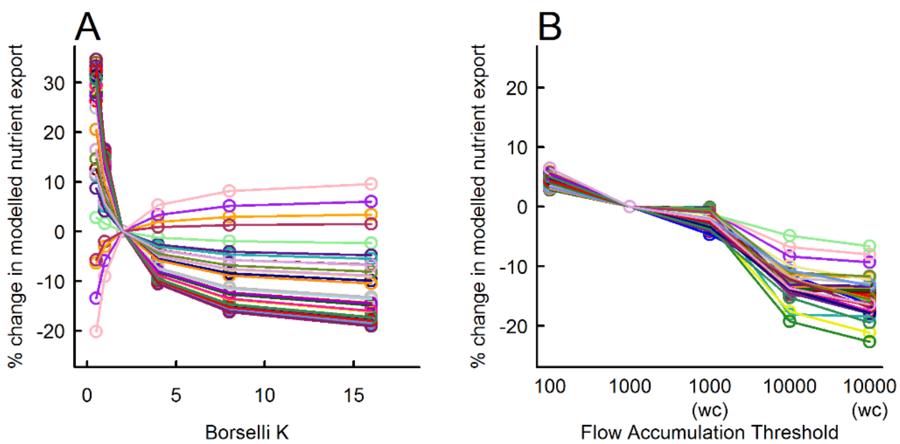

Fig. 2 Sensitivity of the NDR model output to variation in the values of A) Borselli k_b parameter and B) Flow accumulation threshold, TFA (with wc indicating where the threshold was applied along with known watercourses from the digital watercourse network being added to the LULC raster). Each colour represents a different catchment.

Sensitivity to variation in the flow accumulation threshold TFA was also catchment specific (Fig 2B). This was unsurprising as the degree to which a given TFA value accurately represents actual watercourses will vary among catchments depending on their hydrogeology and topography. As can be seen in Figure 4, the default value of 1000 overestimated the stream density in some catchments whilst underestimating it in others. Thus, either reducing or increasing the threshold improved representation of the routing of nutrients in some catchments but made it less accurate in others – values of 100 captured most watercourses in some catchments (Fig. 3A and 3B) whilst in others actual watercourses were best represented by TFA of 10,000 (Fig. 3C and 3D). Addition of mapped watercourses to the LULC input with a TFA of 1000 resulted in comparatively minor changes to the nutrient export (Fig 2B), but ensured that no catchment had known watercourses which were not modelled as such. Using the same approach with a TFA of 10,000 had a large effect on the modelled nutrient export (Fig 2B), reducing nutrient export by up to 20%, by restricting in-stream transport to mapped watercourses only. Which of these latter results is the more accurate is likely to depend on the accuracy of the mapped watercourse network (Baker et al. 2007), many of which, for example ditches and field drains, have not been mapped into a hydrologically consistent network for the UK. Because small, unmapped watercourses are known to have a potentially high impact on nutrient flux (Edwards and Withers 2008; Foster et al. 2003; Heathwaite et al. 2006) we chose to use a TFA value of 1000 with watercourses from Moore et al. (1994) added to the LULC raster for further analyses.

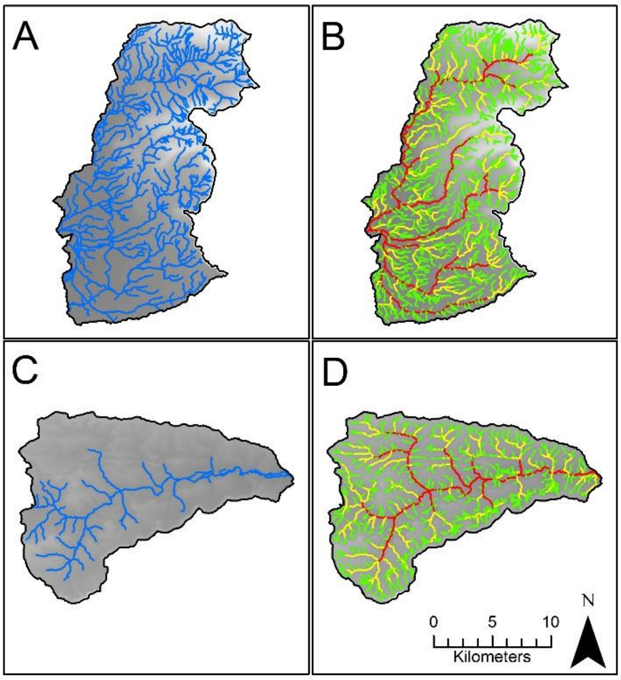

Fig. 3 Examples of two catchments showing the catchment specific effects of variation in the flow accumulation threshold, TFA, on modelled watercourse location. Panels A and C show the known watercourse network (in blue) overlain onto the hydrologically corrected digital elevation model. Panels B and D show streams as determined by three flow accumulation thresholds (100 = green;

1000 = green + yellow; 10,000 = green + yellow + red). The catchments are shaded according to altitude from dark (low, minimum = sea level) to pale (high, maximum = 600 m.a.s.l) grey.

3.2. Model validation and comparison of input datasets

Whilst the slope of the relationship remained similar for both nutrients, both N and P showed increasing percentage differences at resolutions coarser than 100m (Table 1 and Figure 4A and D). When reporting percentage differences across catchments we used the median of mathematical absolute percentage differences to avoid spurious impressions of increased average accuracy resulting from a wider range of under- and overestimates. At coarser (>100m) resolutions, although absolute values became increasingly erroneous for both nutrients, modelled N tended to preserve relative magnitudes of differences between catchments (shown by slightly increased Spearman's ρ). Indeed, the relatively stable values for r_{LR}^2 for N suggest that coarser resolutions gave increasingly severe underestimates, but that the relationship between modelled and measured data remained relatively consistent across catchments. In contrast, at coarser resolutions than 100m, modelled P became increasingly inaccurate in terms of both absolute and relative export, and the relationship between modelled and measured data became increasingly inconsistent (table 1).

In practical terms, finer resolutions substantially increased the model run time, from around 30 seconds at 800m resolution, through 5 minutes at 100m resolution to around 4 hours at 25m resolution. The size of the input and output files was also substantially greater at finer resolutions, with output export maps for a single nutrient of 1.5 gigabytes, 100 megabytes and 2 megabytes for resolutions of 25, 100 and 800 metres, respectively. Given the observed drop off in r_{LR}^2 and Spearman's ρ for P and the increased percentage difference between modelled and measured data for both nutrients at resolutions coarser than 100m (Table 1 and Figure 4) we selected a resolution of 100m for further model testing and validation.

Table 1 Comparisons of total P and N export from the InVEST NDR model with exports estimated from measured flows and nutrient concentrations, for varying resolutions of input data. Estimated exports were adjusted to remove point sources. Results are: median absolute percentage difference; Spearman's ρ and the intercept, slope and r^2 (r_{LR}^2) of a linear regression; between the two datasets.

Nutrient	Resolution (m)	Median absolute % difference	Spearman's rho (ρ)	Linear regression		
				Intercept	Slope (± 95% CI)	r_{LR}^2
Phosphorus	25	54.51	0.77	0.31	0.49 (±0.12)	0.72
	50	56.43	0.78	0.34	0.49 (±0.12)	0.71
	100	55.73	0.79	0.34	0.49 (±0.12)	0.73

	200	56.30	0.79	0.31	0.48 (±0.13)	0.69
	400	67.91	0.75	0.15	0.47 (±0.14)	0.62
	800	88.96	0.56	-0.28	0.44 (±0.23)	0.36
Nitrogen	25	72.57	0.75	0.31	0.67 (±0.27)	0.71
	50	70.37	0.78	0.33	0.67 (±0.27)	0.72
	100	72.58	0.81	0.28	0.69 (±0.25)	0.76
	200	76.56	0.83	0.15	0.71 (±0.23)	0.80
	400	84.11	0.87	-0.25	0.79 (±0.23)	0.81
	800	95.51	0.88	-1.28	0.98 (±0.37)	0.73

Because the sensitivity of the model to k_b appeared relatively high (Fig. 3A), and because there was no clear way to assess which value was most appropriate from our sensitivity analysis alone, we ran the model and compared to validation data for values of k_b of 0.5, 1, 1.5, 2, 2.5, 3, 3.5 and 4. We did not explore values of k_b beyond the range 0.5 - 4 here because sensitivity analysis demonstrates that at values much over 4 the impact of Kb on the model levels off, whilst at values approaching zero, the results diverge towards extreme values (Figure 2A). Overall, the effect of varying k_b on the fit to the validation data was not large, with near identical r_{LR}^2, slope and Spearman's correlation coefficient (Table 2). From Figure 4B and 4E, it can be seen that lower k_b values resulted in median percentage differences closer to zero, but this appears due to an increased number of outliers with substantial overestimates rather than a general improvement across catchments. This is perhaps unsurprising, given the widely varying catchment responses to changes in k_b seen in Figure 3A. There was thus no clear evidence to support altering the value of k_b from the default of 2 for our modelling across multiple catchments.

Table 2 Comparisons of P and N export from the InVEST NDR model with exports estimated from measured flows and nutrient concentrations (adjusted to remove point sources), for eight values of k_b.

Nutrient	k_b	Median absolute % difference	Spearman's rho (ρ)	Linear regression		
				Intercept	Slope (± 95% CI)	r_{LR}^2
Phosphorus	0.5	41.16	0.77	0.41	0.49 (±0.12)	0.71
	1	53.97	0.76	0.37	0.49 (±0.12)	0.71
	1.5	58.43	0.77	0.33	0.49 (±0.12)	0.71
	2	55.73	0.79	0.34	0.49 (±0.12)	0.73
	2.5	56.99	0.79	0.32	0.49 (±0.12)	0.72
	3	55.41	0.79	0.31	0.49 (±0.12)	0.72
	3.5	53.59	0.79	0.30	0.49 (±0.12)	0.72

	4	54.54	0.79	0.29	0.49 (±0.12)	0.72
Nitrogen	0.5	64.00	0.78	0.38	0.68 (±0.24)	0.75
	1	72.99	0.78	0.32	0.68 (±0.24)	0.75
	1.5	75.49	0.80	0.27	0.68 (±0.24)	0.76
	2	72.58	0.81	0.28	0.69 (±0.25)	0.76
	2.5	73.72	0.81	0.25	0.69 (±0.25)	0.76
	3	74.52	0.81	0.22	0.70 (±0.25)	0.76
	3.5	75.11	0.81	0.21	0.70 (±0.25)	0.76
	4	75.56	0.81	0.19	0.70 (±0.24)	0.77

Having explored the effect of k_b and the input data resolution, we then compared the three input precipitation data sources. The choice of precipitation data again made comparatively little difference to either N or P export (Table 3 and Figure 4C and 4F). The randomised precipitation dataset did show reductions in ρ and r_{LR}^2 but actually decreased median percentage difference.

Table 3 Comparisons of total P and N export from the InVEST NDR model with exports estimated from measured flows and nutrient concentrations (adjusted to remove point sources), for three difference sources of precipitation data (WorldClim, Met Office UKCP09 and CEH-GEAR).

Nutrient	Precipitation data source	Median absolute % difference	Spearman's rho (ρ)	Linear regression		
				Intercept	Slope (± 95% CI)	r_{LR}^2
Phosphorus	WorldClim	56.40	0.81	0.33	0.51 (±0.12)	0.73
	UKCP09	55.73	0.79	0.34	0.49 (±0.12)	0.73
	CEH-GEAR	57.07	0.77	0.35	0.49 (±0.12)	0.71
	Random	55.13	0.69	0.37	0.46 (±0.17)	0.53
Nitrogen	WorldClim	70.70	0.88	0.17	0.74 (±0.21)	0.83
	UKCP09	72.58	0.81	0.28	0.69 (±0.25)	0.76
	CEH-GEAR	73.59	0.84	0.28	0.69 (±0.25)	0.75
	Random	65.27	0.74	0.28	0.68 (±0.27)	0.71

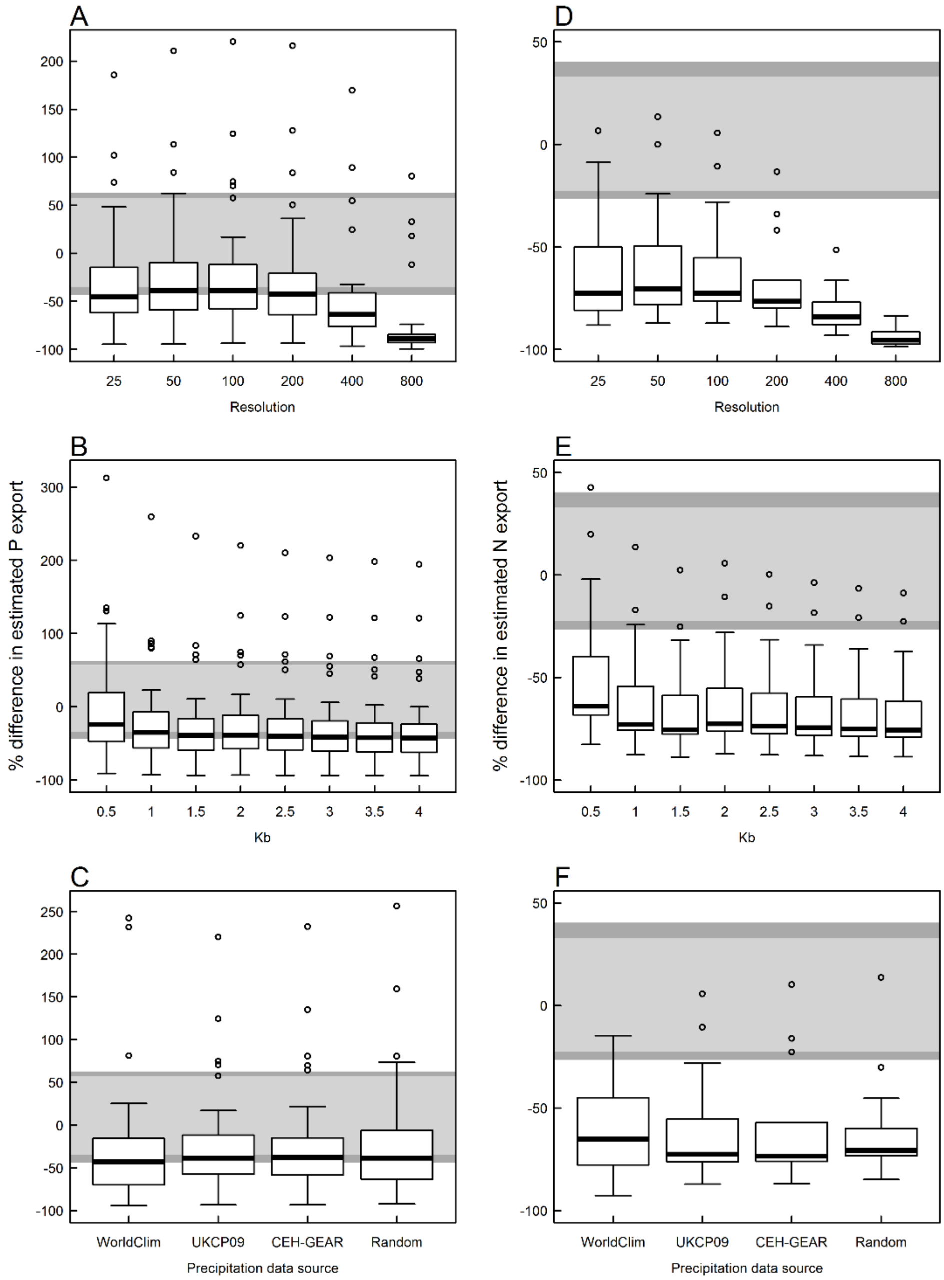

Fig. 4 Boxplots showing the effect of spatial resolution (i.e. dimensions of raster cells in metres) (A,D), Borselli k_b (B,E) and precipitation data source (C,F) on percentage differences between estimated total nutrient export per catchment from the InVEST NDR model and corresponding exports estimated from gauged flow and measured nutrient concentration data (adjusted to remove point sources), for phosphorus (A,B,C) and nitrogen (D,E,F). Grey shaded areas indicate the range of variation in estimated nutrient export values resulting from interannual variation in estimated exports (quartiles, light grey) and the maximum and minimum values for average per capita nutrient outflow from point sources (dark grey)

Modelled total nutrient export showed a better fit to the empirical data than did modelled load alone (P: r_{LR}^2 = 0.73, 0.56, Spearman's ρ = 0.79, 0.69, N: r_{LR}^2 = 0.76, 0.72, Spearman's ρ = 0.80, 0.75, for load and export, respectively, with 100m resolution inputs, K_b = 2 and UKCP09 precipitation data). The NDR factor component of the model thus results in substantial increases in model performance over a simple summation of loads, especially for P.

Because the results at all values of k_b and the different precipitation datasets resulted in good predictions of the relative magnitude of nutrient export (ρ = 0.77 - 0.81 and 0.75 - 0.88, for phosphorous and nitrogen, respectively) but relatively large underestimates of absolute values (range of absolute median estimates ±44.4% – ±58.4% and ±65.3% – ±76.6% for phosphorous and nitrogen, respectively), we ran a final model with reduced retention coefficients for both nutrients. Whilst this deviates from parameter values reported from empirical studies (see section 2.2), we were interested to see if a large improvement in accuracy could be made by performing a simple, uniformly applied adjustment to retention values. We therefore divided retention values by two and re-ran the model (with 100m resolution inputs, K_b = 2 and UKCP09 precipitation data). Although this resulted in slightly reduced absolute median percentage differences (by 8.5% and 9.2% for phosphorous and nitrogen, respectively), the Spearman's ρ and the slope and r_{LR}^2 from linear regression were also reduced (4-10% reduction Spearman's ρ, 4-13% reduction in r_{LR}^2, 8%-12% reduction in slope). This suggests that modifying the retention coefficients away from literature values helps to reduce the median level of underestimates, but reduces the ability of the model to predict relative magnitude of nutrient export between catchments, probably by worsening overestimation in low exporting catchments whilst improving underestimation in high yielding ones (Figure 5).

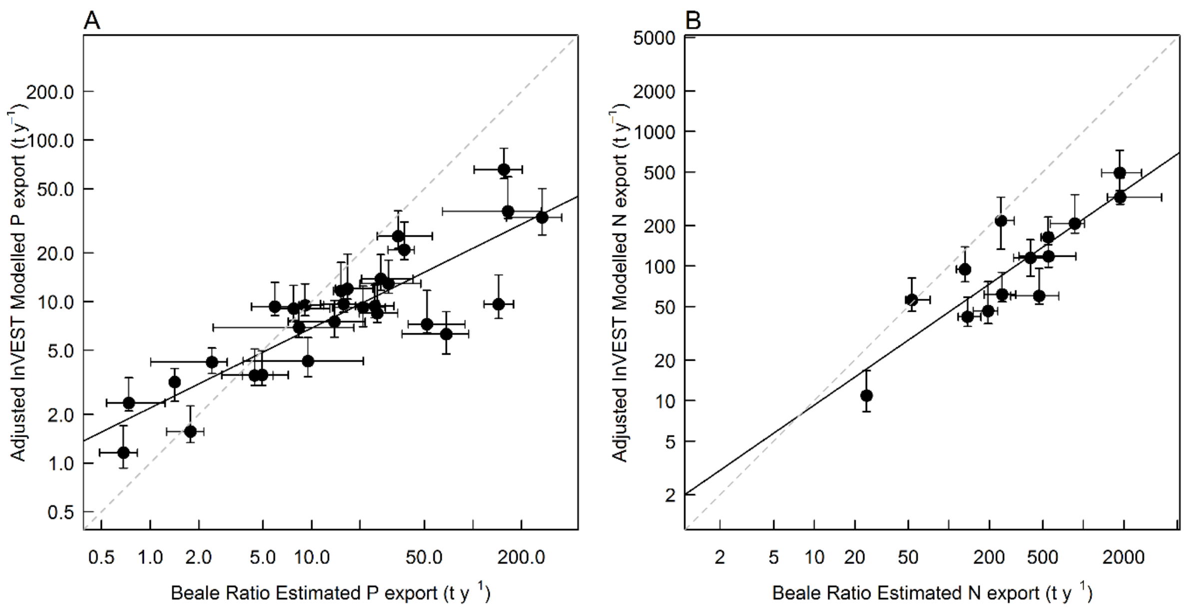

Fig. 5. Nutrient export per catchment from the InVEST NDR model plotted against exports estimated from measured flows and nutrient concentrations (adjusted to remove point sources), for P (panel A) and N (panel B). Points represent InVEST results (input resolution = 100m, k_b = 2, precipitation data = UKCP09) against Beale Ratio Estimated nutrient export from measurements. Horizontal bars span the range given by 25th to 75th percentile of interannual variation in the Beale ratio estimated nutrient export ± the maximum and minimum values for average *per capita* nutrient outflow from point sources. Vertical bars indicate the range in modelled export resulting from running the model with values of k_b between 0.5 and 4, input raster resolution of 25, 50, 100 and 200 metres and the three different precipitation datasets. A 1:1 relationship is indicated by the dotted line. Note axes are on a log10 scale.

4. Discussion

4.1. Performance of the InVEST NDR model

Our results suggest that the InVEST NDR model can give good results in terms of the relative magnitude of N and P export across a wide variety of UK river catchments, with ρ between 0.7 and 0.83 depending on the scale of input data and parameter values used. However, accuracy in terms of estimating actual nutrient export was comparatively low with the majority of catchments showing over or underestimates of up to 44% for P and 65% for N. It should be noted that attempting to gain good model performance over a large number of widely varying catchments is a challenging test for the model. Performance is expected to be higher with calibration at the regional level with catchments having similar hydrogeological properties. Whilst some studies perform such model performance assessment (e.g. Bai et al. 2013; Terrado et al. 2014), many ecosystem service models are applied at regional or national scales without validation (Martínez-Harms and Balvanera 2012). A survey across sub-Saharan Africa demonstrated that many stakeholders wish to run ecosystem service models at national scales (Willcock et al. 2016). Furthermore, ecosystem service models are often perceived as being of great use in data-scarce parts of the world (Pandeya et al. 2016; Villa et al. 2014) where there are few opportunities to calibrate or validate. Therefore, it is important for studies such as ours to demonstrate some of the possible pitfalls of applying ecosystem service models without extensive validation and sensitivity testing.

4.2. Understanding model sensitivities

Sensitivity to variation in the input parameter values is unsurprising and, of course, desirable if a model is to be used to assess change over time or among future change scenarios. However, it is also important to understand that such sensitivities can determine how appropriate a model is to a particular study region, where to focus most effort on data acquisition (Boithias et al. 2014; Sánchez-

Canales et al. 2012), or to aid in assessing the uncertainty associated with model outcomes. In brief, the model appeared most sensitive to the nutrient loading and retention values, the threshold flow accumulation and the resolution of the input raster data (beyond a certain range). We discuss each of the parameters in turn.

4.2.1. *Nutrient load and retention*

The linear response between nutrient export and nutrient load and nutrient retention coefficients is to be expected, given that nutrient export is calculated as the product of nutrient load on a pixel and the NDR factor, which is proportional to nutrient retention parameters from downslope pixels. These parameters are thus the major drivers by which the spatial configuration of land use/land cover affects nutrient runoff. Importantly, nutrient loads and retention efficiencies will vary greatly in time and space. In our test catchments, most of which are dominated by arable land or agriculturally-improved grassland, such variation will be driven by crop type, stocking density, fertiliser application rates and timings, and other farm management practices. It is, therefore, essential to research these values sufficiently to ensure that they are robust for the land cover types that are dominant in the study region and those that are of most interest in relation to any change scenarios that are being explored.

4.2.2. *k_b parameter*

The Borselli k_b parameter determines the relationship between hydrologic connectivity (the degree of connection from patches of land to the stream) and the NDR. Higher values mean that the relationship between the connectivity index and the NDR factor becomes linear, whereas lower values mean that this relationship becomes a step function. This relationship is site-specific, as demonstrated by the very different responses to varying k_b shown by different catchments in our sensitivity analysis. This is also likely to be the reason that, from our results, calibration to produce the best cross-catchment absolute accuracy may not result in the most accurate predictions of relative magnitude between catchments and *vice versa*. Therefore, although this parameter is in practice the main parameter used for calibration (Sharp et al. 2016), where possible k_b should be determined regionally, across catchments with similar hydrogeological properties.

4.2.3. *Threshold flow accumulation*

Varying the flow accumulation threshold TFA had a substantial effect on model output. This effect is partly explained by the model structure, which assumes that stream pixels do not export any nutrient. Therefore, changing the density of the stream network also changes the number of pixels that actually contribute to nutrient loading and retention (e.g. 66%, 92%, 98%, 99% at TFAs of 10, 100, 1000 and 10000, respectively, at 25m DEM resolution). Our results show that, as with k_b, selecting a single value that is equally applicable across a number of catchments is difficult, because catchment topography and hydrogeological attributes (e.g. groundwater flow) can change the threshold that needs to be set to capture actual watercourses. Comparing the derived stream network to a known watercourse network is a key first step to selecting an appropriate value, and our results also suggest that modifying the DEM and LULC map to capture known watercourse networks may provide a robust approach to overcoming this issue when conducting cross-catchment analyses.

4.2.4. *DEM and LULC raster resolution*

Changing the resolution of the input DEM and LULC spatial data had comparatively little effect on the accuracy of the model output for both P and N at resolutions less than or equal to 100m. Whilst this is in contrast to other studies which have concluded that increased data resolution usually results in increased model accuracy (Brazier et al. 2005), decreased sensitivity to input raster resolution is a stated aim of the design of the NDR model (Sharp et al. 2016), hence the inclusion of TFA and critical flow length parameters which the user can modify. It appears that resolutions finer than 100m gain little in absolute accuracy to justify the very substantial increases in file size (making data harder to store, manage and disseminate) and running time which result from running the model with finer resolution inputs.

However, resolutions coarser than 100m resulted in decreasing accuracy, especially for P. This is likely to be a result of coarser resolution cells losing spatial detail, with values being generalised to average (DEM) or dominant (LULC) values per cell. The most likely mechanism for the effects we observed are loss of detail from the LULC raster. If the key LULC classes governing nutrient export are relatively small in area, they may be lost from aggregated inputs. For example, in UK upland catchments which are largely semi-natural, small areas of agricultural land close to watercourses would have a disproportionate effect on total nutrient export, but may not form the majority cover of any non-watercourse pixels in a coarse resolution LULC map, removing their potential to influence modelled nutrient export. The two nutrients differed somewhat in their responses to resolution (with N retaining accurate relative magnitude and a consistent relationship between modelled and measured data, even though underestimation became more severe). This is probably because of their different loadings and export pathways. Phosphorus is more associated with high releases from proportionally small areas with high hydrologic connectivity whilst nitrogen is more evenly spread across land cover classes and less directly linked to the degree of hydrologic connectivity (Edwards and Withers 2008; Withers and Lord 2002), such that the loss of spatial detail at coarser resolutions affects the ability of the model to reflect actual export to different degrees.

4.2.5 Precipitation data source

Unlike the InVEST water yield model (Redhead et al. 2016), the NDR model appeared relatively insensitive to the source of input precipitation data. All three datasets produced similar results, and even the randomised data only reduced accuracy slightly. To some extent this is unsurprising. The effect of precipitation data is to modify the per pixel load to account for runoff potential by relating the precipitation per cell to the average across the raster (see Supplementary Material, Appendix S1). Therefore, providing that general spatial patterns are preserved between input datasets, this should be sufficient to obtain similar results. The lack of effect of using randomised data is perhaps more surprising, as here the spatial pattern of relative runoff has been removed. However, by using long term average data at 1km to 5km scales, the range of values is not high within many catchments, so even when randomising the data the distribution of runoff potential across the landscape does not vary hugely (Supplementary Material, Table S2). Of course, for those catchments with a higher range in precipitation (in our analysis this was limited to larger catchments spanning upland and lowland), randomisation will have a greater effect, so in locations where rainfall is more variable within catchments (e.g. Boithias et al. 2014; Terrado et al. 2014), or over timescales where temporal variation becomes an issue, this parameter may become of much greater importance.

4.3. LIMITATIONS OF THE MODEL

The InVEST NDR model includes only a relatively limited number of the wide range of complex, and spatially and temporally variable processes that influence nutrient transport from land to watercourses (see reviews in Arheimer and Lidén 2000; Edwards and Withers 2008; Parn et al. 2012). Whilst this is clearly stated in the InVEST documentation, it is important to explore some of these limitations to remind potential users of the sensible use of the model and to explain the relatively large and variable underestimates of nutrient delivery that our results show.

One of the most obvious limitations of applying this model within the UK is that it focuses on diffuse (i.e. non-point) sources of nutrient only, while most UK catchments, especially those in more populated areas, are also affected by nutrient discharges from WWTWs. This is not a limitation of the model as such, but it is a problem that needs to be addressed when comparing modelled output with measured values. This is discussed below under limitations of our validation approach (Section 4.3).

A limitation of the model that is harder to compensate for is the presence of catchment specific processes that may affect nutrient transport and export in ways that are hard to predict or capture within model frameworks that are based on an average load per area of land use/land cover class. These include nutrient releases from so-called intermediate sources (because they are neither diffuse nor a predictable point source) such as field drains, septic tanks, farmyard and/or road/track runoff (Edwards and Withers 2008). Such features are difficult to include as a LULC class because they are rarely well mapped and nutrient releases from them are often difficult to predict because of high spatial and temporal variation (Edwards and Withers 2008; Withers et al. 2014). For example, field drains can release large amounts of P into watercourses from agricultural land during storm events, bypassing surface flow and normal retention capabilities (Foster et al. 2003; Heathwaite et al. 2006; Hooda et al. 1999). Such features may be especially important in rural catchments where most other sources are diffuse (Jarvie et al. 2003). In addition, it has been shown that interpolation of infrequent data is unlikely to give reliable estimates of in-stream P loads where temporal changes in stream flow and P concentrations happen very quickly in response to rainfall and surface runoff (Defew et al. 2013).

The model can be set to apportion a set amount of nutrient transport to subsurface flow for each LULC class; this is then subject to a simple exponential decay function driven by distance to stream. A value can be defined by the user across all LULC classes (Sharp et al. 2016), but in reality subsurface flow and nutrient retention varies considerably within LULC classes. There are also many features that, whilst contributing to nutrient retention and export, lie below the spatial resolution of most input LULC maps. These include riparian buffer strips or riparian vegetation that can retard or reduce the level of nutrients entering the watercourse (Aguiar Jr et al. 2015; Darch et al. 2015; Lena et al. 1994; Parn et al. 2012). Once nutrient enters a watercourse it may be subject to further retention by aquatic vegetation or uptake by riverine sediments (Jarvie et al. 2005). However, on an annual scale, most of these in stream nutrient sinks are temporary and much of the nutrient delivered to a watercourse from land eventually leaves the catchment in one form or another (Bowes and House 2001).

Although the two nutrients are modelled in identical ways by the InVEST model, the extent to which the model is able to reflect the real world flow of the two nutrients is likely to differ, hence our observed differing accuracies for N and P. This is because of differences in anthropogenic sources, temporal and spatial variation is levels of output, and the chemical properties of the two elements and the various forms in which they are usually transported through soil-water systems. One key difference is that N can be removed from the hydrological system by denitrification to atmospheric N_2 and, in some cases, very high retention can be achieved within a watercourse by riverine or wetland vegetation that promotes such processes (Parn et al. 2012; Saunders and Kalff 2001). No equivalent process exists for P (Parn et al. 2012), so at times of high P runoff, the normal retention capacity of any particular land cover class may be more likely to become saturated, leading to higher than expected exports (Koerselman et al. 1990). Phosphorus flows are often dominated by point source releases and temporal factors such as surface runoff during and after storm events. In contrast, N transport is more often associated with broader land cover patterns, subsurface flow and soil chemistry (Edwards and Withers 2008; Nedwell et al. 2002; Parn et al. 2012; Withers and Lord 2002).

The issues outlined above may be part of the reason why a simple, universally applied reduction of retention coefficients did not substantially improve model accuracy. However, it is also worth noting that the ability of the model to obtain good predictions in terms of the relative magnitude of nutrient export, despite these limitations, suggests that the model and its results are useful if interpreted with caution, especially in order to identify spatial patterns of N or P delivery across catchments or to examine relative change under potential scenarios, which is the intended use for most InVEST models (Sharp et al. 2016). However, the relative export or retention of nutrients alone may not be sufficiently informative for decision makers, who may need to know whether export is sufficient to meet a threshold (e.g. a legal maximum for drinking water or a level known to cause certain ecological impacts) or to place an economic value the service of nutrient retention in terms of avoided water treatment costs. In this case, an understanding of the absolute accuracy of modelled nutrient export figures, and how to best improve this, is key. Of note, the model is open-source and its code is regularly updated by the development team or external contributors so that such limitations may be addressed in the future. For example, the NDR model used here was already an improvement over a previous version (Water Quality model, InVEST v3.2).

4.4. Limitations of the validation approach

Validation of the model using the approach detailed in this paper has its limitations. Without actual measurements of nutrient export to water, estimations of average annual export will always be subject to a degree of error arising from a variety of factors whatever the method of calculation used.

Firstly, whilst the Beale ratio approach to calculating nutrient has been shown to provide better results than other methods (Dolan et al. 1981; Dunn et al. 2014; Meals et al. 2013; Quilbé et al. 2006; Richards and Holloway 1987), it has the potential to underestimate in-stream nutrient load if nutrient sampling does not coincide with periods of peak flow (Quilbé et al. 2006) or peak runoff, as may occur during short duration, extreme weather events. During such events, P transport is very difficult to measure accurately unless sampled at very high frequencies, which is rarely the case for routine monitoring data (Defew et al. 2013). Also, the peak flows recorded by gauging stations may themselves be underestimates where these events affect the accurate measurement of flow (e.g. bypassing of the gauging station by groundwater or flooding, water transfer, etc.). However, our results suggested that BRE derived values were mostly larger than the modelled N and P values, even when compared to the interquartile range of BRE values across years or the inter-annual ranges per catchment, so this is unlikely to be a major driver of this apparent error in model predictions.

Because the model only accounts for nutrients derived from surface runoff, it was necessary to adjust the validation data to estimate the total that would be derived from diffuse sources, only. Using WWTW locations and average per capita nutrient export values is common practice, but potential per capita figures show wide variation between studies, catchments and over time (Edwards and Withers 2008; Johnes 1996; Naden et al. 2016). However, this variation is unlikely to show a systematic bias towards over- or under-estimation across catchments and so our results should provide a fair reflection of model performance in terms of the slope of the linear regression line, even if individual catchments over- or under-estimate the proportion of nutrient export that is derived from point sources. We also quantified the likely extent of this potential error by examining the variation in estimated diffuse source nutrient export imparted by varying the maximum and minimum per capita values for nutrient export from point sources. Even so, there remains a potential for unquantified error in terms of unmapped point sources and variation in per capita values among catchments. Because we excluded catchments where point source nutrient exports appeared to contribute over 50% to the total in-stream nutrient load, we also excluded heavily urbanised catchments. So, our validation cannot inform on the ability of the model to predict diffuse pollution in these types of catchment.

4.5. Conclusions

Whilst the InVEST NDR model gives good estimates of the relative magnitude of nutrient exports across catchments, absolute values are frequently underestimated even after calibration of input parameter values. This is to be expected given the simple nature of the InVEST model and the aims of using it to compare the outcomes of change scenarios across a wide range of ecosystem services (Sharp et al. 2015). Key model sensitivities were to nutrient loading and retention factors and the threshold flow accumulation. Input raster resolution had major impacts on model performance only at resolutions coarser than 100m. For resolutions finer than this, there was little in the way of increased accuracy to offset the increased model run time and output data volume.

Collating the data sources for input and validation of the model, even in such a well-studied region such as the UK, was time consuming and complex. Similar difficulties are likely to be encountered in regions that have less frequent monitoring schemes for nutrients and water flow. Since one of the stated aims of the InVEST model is to allow meaningful analyses to take place in data-poor regions, we recommend the following uncertainty assessment analyses: exploration of alternative input datasets for the study region, sensitivity analyses on loads and retention efficiencies for dominant LULC types, TFA, and k_b, and a thorough exploration of the model outputs before using them to inform decisions. This reflects the recommendations of the designers of the InVEST NDR model (Hamel et al. 2015; Sharp et al. 2016) and the findings of previous studies across a number of ecosystem services (Boithias et al. 2014; Pessacg et al. 2015; Redhead et al. 2016; Sánchez-Canales et al. 2012).

Acknowledgements

Thanks to Matthew Fry for advice and assistance on analysing NRFA data, Olivia Hitt for assistance with extracting and handling WIMS water chemistry data, Virginie Keller for providing point source data and Justyna Olszewska for assistance with researching loading and export coefficients from the literature. This work was performed using National Capability funding from the Natural Environmental Research Council under project NEC04936.

References

Aguiar Jr, T.R., Rasera, K., Parron, L.M., Brito, A.G., Ferreira, M.T., 2015. Nutrient removal effectiveness by riparian buffer zones in rural temperate watersheds: The impact of no-till crops practices. Agricultural Water Management 149, 74-80.

Arheimer, B., Lidén, R., 2000. Nitrogen and phosphorus concentrations from agricultural catchments—influence of spatial and temporal variables. Journal of Hydrology 227, 140-159.

Arnold, J.G., Srinivasan, R., Muttiah, R.S., Williams, J.R., 1998. Large area hydrologic modeling and assessment part i: model development. JAWRA Journal of the American Water Resources Association 34, 73-89.

Bai, Y., Zheng, H., Ouyang, Z., Zhuang, C., Jiang, B., 2013. Modeling hydrological ecosystem services and tradeoffs: a case study in Baiyangdian watershed, China. Environmental Earth Sciences 70, 709-718.

Baker, M.E., Weller, D.E., Jordan, T.E., 2007. Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landscape Ecology 22, 973-992.

Beale, E.M.L., 1962. Some uses of computers in operational research. Industrielle Organisation 2, 51-52.

Boithias, L., Acuña, V., Vergoñós, L., Ziv, G., Marcé, R., Sabater, S., 2014. Assessment of the water supply:demand ratios in a Mediterranean basin under different global change scenarios and mitigation alternatives. Science of the total environment 470–471, 567-577.

Bowes, M., House, W., 2001. Phosphorus and dissolved silicon dynamics in the River Swale catchment, UK: a mass-balance approach. Hydrological Processes 15, 261-280.

Bowes, M.J., Hilton, J., Irons, G.P., Hornby, D.D., 2005. The relative contribution of sewage and diffuse phosphorus sources in the River Avon catchment, southern England: Implications for nutrient management. Science of the total environment 344, 67-81.

Braat, L.C., de Groot, R., 2012. The ecosystem services agenda:bridging the worlds of natural science and economics, conservation and development, and public and private policy. Ecosystem Services 1, 4-15.

Brazier, R.E., Heathwaite, A.L., Liu, S., 2005. Scaling issues relating to phosphorus transfer from land to water in agricultural catchments. Journal of Hydrology 304, 330-342.

Breuer, L., Vache, K., Julich, S., Frede, H.-G., 2008. Current concepts in nitrogen dynamics for mesoscale catchments. Hydrological Sciences Journal 53, 1059-1074.
Darch, T., Carswell, A., Blackwell, M.S.A., Hawkins, J.M.B., Haygarth, P.M., Chadwick, D., 2015. Dissolved Phosphorus Retention in Buffer Strips: Influence of Slope and Soil Type. Journal of Environmental Quality 44, 1216-1224.
Defew, L.H., May, L., Heal, K.V., 2013. Uncertainties in estimated phosphorus loads as a function of different sampling frequencies and common calculation methods. Marine and Freshwater Research 64, 373-386.
Dennedy-Frank, P.J., Muenich, R.L., Chaubey, I., Ziv, G., 2016. Comparing two tools for ecosystem service assessments regarding water resources decisions. Journal of Environmental Management 177, 331-340.
Dillon, P.J., Kirchner, W.B., 1975. The effects of geology and land use on the export of phosphorus from watersheds. Water Research 9, 135-148.
Dolan, D.M., Yui, A.K., Geist, R.D., 1981. Evaluation of River Load Estimation Methods for Total Phosphorus. Journal of Great Lakes Research 7, 207-214.
Drewry, J.J., Newham, L.T.H., Greene, R.S.B., 2011. Index models to evaluate the risk of phosphorus and nitrogen loss at catchment scales. Journal of Environmental Management 92, 639-649.
Dunn, S.M., Sample, J., Potts, J., Abel, C., Cook, Y., Taylor, C., Vinten, A.J.A., 2014. Recent trends in water quality in an agricultural catchment in Eastern Scotland: elucidating the roles of hydrology and land use. Environmental Science: Processes & Impacts 16, 1659-1675.
Edwards, A.C., Withers, P.J.A., 2008. Transport and delivery of suspended solids, nitrogen and phosphorus from various sources to freshwaters in the UK. Journal of Hydrology 350, 144-153.
Envrionment Agency, 2017. Water quality data archive.
Feng, M., Liu, S., Euliss Jr, N.H., Young, C., Mushet, D.M., 2011. Prototyping an online wetland ecosystem services model using open model sharing standards. Environmental Modelling & Software 26, 458-468.
Fisher, B., Turner, R.K., Morling, P., 2009. Defining and classifying ecosystem services for decision making. Ecological economics 68, 643-653.
Foster, I.D., Chapman, A., Hodgkinson, R., Jones, A., Lees, J., Turner, S., Scott, M., 2003. Changing suspended sediment and particulate phosphorus loads and pathways in underdrained lowland agricultural catchments; Herefordshire and Worcestershire, UK, In The Interactions between Sediments and Water. pp. 119-126. Springer.
Fozzard, I., Doughty, R., Ferrier, R.C., Leatherland, T., Owen, R., 1999. A quality classification for management of Scottish standing waters. Hydrobiologia 395, 433-454.
Fry, M.J., Swain, O., 2010 Hydrological data management systems within a national river flow archive, In Role of Hydrology in Managing Consequences of a Changing Global Environment. ed. C. Kirby, pp. 808-815. British Hydrological Society.
Grafius, D.R., Corstanje, R., Warren, P.H., Evans, K.L., Hancock, S., Harris, J.A., 2016. The impact of land use/land cover scale on modelling urban ecosystem services. Landscape Ecology 31, 1509-1522.
Hamel, P., Chaplin-Kramer, R., Sim, S., Mueller, C., 2015. A new approach to modeling the sediment retention service (InVEST 3.0): Case study of the Cape Fear catchment, North Carolina, USA. Science of the total environment 524, 166-177.
Hamel, P., Falinski, K., Sharp, R., Auerbach, D.A., Sánchez-Canales, M., Dennedy-Frank, P.J., 2017. Sediment delivery modeling in practice: Comparing the effects of watershed characteristics and data resolution across hydroclimatic regions. Science of the total environment 580, 1381-1388.
Hamel, P., Guswa, A.J., 2015. Uncertainty analysis of a spatially explicit annual water-balance model: case study of the Cape Fear basin, North Carolina. Hydrological Earth System Science 19, 839-853.
Hamel, P., Sharp, R., 2017. InVEST 3.3.3 Nutrient Delivery Ratio Model, Zenodo.
Heathwaite, A.L., Burke, S.P., Bolton, L., 2006. Field drains as a route of rapid nutrient export from agricultural land receiving biosolids. Science of the total environment 365, 33-46.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A., 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25, 1965-1978.
Hooda, P.S., Moynagh, M., Svoboda, I.F., Edwards, A.C., Anderson, H.A., Sym, G., 1999. Phosphorus Loss in Drainflow from Intensively Managed Grassland Soils. Journal of Environmental Quality 28, 1235-1242.
Jackson, D.L., 2000. Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationship with other classifications.
Jarvie, H.P., Jürgens, M.D., Williams, R.J., Neal, C., Davies, J.J., Barrett, C., White, J., 2005. Role of river bed sediments as sources and sinks of phosphorus across two major eutrophic UK river basins: the Hampshire Avon and Herefordshire Wye. Journal of Hydrology 304, 51-74.
Jarvie, H.P., Neal, C., Withers, P.J.A., Robinson, A., Salter, N., 2003. Nutrient water quality of the Wye catchment, UK: exploring patterns and fluxes using the Environment Agency data archives. Hydrol. Earth Syst. Sci. 7, 722-743.
Jenkins, G.J., Perry, M.C., Prior, M.J., 2008. The climate of the United Kingdom and recent trends. Hadley Centre, Exeter, UK.
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. Freshwater Biology 36, 451-473.
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, 323-349.
Keeler, B.L., Polasky, S., Brauman, K.A., Johnson, K.A., Finlay, J.C., O'Neill, A., Kovacs, K., Dalzell, B., 2012. Linking water quality and well-being for improved assessment and valuation of ecosystem services. Proceedings of the National Academy of Sciences 109, 18619-18624.
Keller, A.A., Fournier, E., Fox, J., 2015. Minimizing impacts of land use change on ecosystem services using multi-criteria heuristic analysis. Journal of Environmental Management 156, 23-30.
Koerselman, W., Bakker, S.A., Blom, M., 1990. Nitrogen, Phosphorus and Potassium Budgets for Two Small Fens Surrounded by Heavily Fertilized Pastures. Journal of Ecology 78, 428-442.
Leh, M.D.K., Matlock, M.D., Cummings, E.C., Nalley, L.L., 2013. Quantifying and mapping multiple ecosystem services change in West Africa. Agriculture, Ecosystems & Environment 165, 6-18.
Lena, B.M.V., Dahl, J., Carsten Lauge, P., Jean, O.L., xe, re, 1994. Nutrient Retention in Riparian Ecotones. Ambio 23, 342-348.
Maes, J., Egoh, B., Willemen, L., Liquete, C., Vihervaara, P., Schägner, J.P., Grizzetti, B., Drakou, E.G., Notte, A.L., Zulian, G., Bouraoui, F., Luisa Paracchini, M., Braat, L., Bidoglio, G., 2012. Mapping ecosystem services for policy support and decision making in the European Union. Ecosystem Services 1, 31-39.
Malinga, R., Gordon, L.J., Jewitt, G., Lindborg, R., 2015. Mapping ecosystem services across scales and continents – A review. Ecosystem Services 13, 57-63.
Martínez-Harms, M.J., Balvanera, P., 2012. Methods for mapping ecosystem service supply: a review. International Journal of Biodiversity Science, Ecosystem Services & Management 8, 17-25.
May, L., House, W.A., Bowes, M., McEvoy, J., 2001. Seasonal export of phosphorus from a lowland catchment: upper River Cherwell in Oxfordshire, England. Science of the total environment 269, 117-130.
May, L., Place, C.J., George, D.G., McEvoy, J., 1996. An Assessment of the Nutrient Loadings from the Catchment to Bassenthwaite Lake, In Report to the Environment Agency. p. 54 pp, North West Region.
McGuckin, S.O., Jordan, C., Smith, R.V., 1999. Deriving phosphorus export coefficients for corine land cover types. Water Science and Technology 39, 47-53.
Meals, D.W., Richards, R.P., Dressing, S.A., 2013. Pollutant load estimation for water quality monitoring projects., In Tech Notes. Developed for U.S. Environmental Protection Agency by Tetra Tech, Inc., Fairfax, VA.

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

Morris, D.G., Flavin, R.W., 1990. A Digital Terrain Model for Hydrology, In Proc 4th Int. Symposium on Spatial Data Handling. pp. 250-262, Zurich.

Morton, D., Rowland, C., Wood, C., Meek, L., Marston, C., Smith, G., Simpson, I.C., 2011. Final report for LCM2007 - the new UK land cover map. , p. 112pp. NERC/Centre for Ecology and Hydrology.

Naden, P., Bell, V., Carnell, E., Tomlinson, S., Dragosits, U., Chaplow, J., May, L., Tipping, E., 2016. Nutrient fluxes from domestic wastewater: a national-scale historical perspective for the UK 1800–2010. Science of the total environment.

Nedwell, D.B., Dong, L.F., Sage, A., Underwood, G.J.C., 2002. Variations of the Nutrients Loads to the Mainland U.K. Estuaries: Correlation with Catchment Areas, Urbanization and Coastal Eutrophication. Estuarine, Coastal and Shelf Science 54, 951-970.

Nelson, E., Mendoza, G., Regetz, J., Polasky, S., Tallis, H., Cameron, D., Chan, K.M.A., Daily, G.C., Goldstein, J., Kareiva, P.M., Lonsdorf, E., Naidoo, R., Ricketts, T.H., Shaw, M., 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. Frontiers in Ecology and the Environment 7, 4-11.

Pandeya, B., Buytaert, W., Zulkafli, Z., Karpouzoglou, T., Mao, F., Hannah, D., 2016. A comparative analysis of ecosystem services valuation approaches for application at the local scale and in data scarce regions. Ecosystem Services 22, 250-259.

Parn, J., Pinay, G., Mander, U., 2012. Indicators of nutrients transport from agricultural catchments under temperate climate: A review. Ecological Indicators 22, 4-15.

Perry, M., Hollis, D., 2005. The generation of monthly gridded datasets for a range of climatic variables over the UK. International Journal of Climatology 25, 1041-1054.

Pessacg, N., Flaherty, S., Brandizi, L., Solman, S., Pascual, M., 2015. Getting water right: A case study in water yield modelling based on precipitation data. Science of the total environment 537, 225-234.

Quilbé, R., Rousseau, A.N., Duchemin, M., Poulin, A., Gangbazo, G., Villeneuve, J.-P., 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: Application to the Beaurivage River (Québec, Canada). Journal of Hydrology 326, 295-310.

R Core Team, 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Redhead, J., Stratford, C., Sharps, K., Jones, L., Ziv, G., Clarke, D., Oliver, T., Bullock, J., 2016. Empirical validation of the InVEST water yield ecosystem service model at a national scale. Science of the total environment 569, 1418-1426.

Richards, R.P., Holloway, J., 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. Water Resources Research 23, 1939-1948.

Sánchez-Canales, M., López Benito, A., Passuello, A., Terrado, M., Ziv, G., Acuña, V., Schuhmacher, M., Elorza, F.J., 2012. Sensitivity analysis of ecosystem service valuation in a Mediterranean watershed. Science of the total environment 440, 140-153.

Saunders, D.L., Kalff, J., 2001. Nitrogen retention in wetlands, lakes and rivers. Hydrobiologia 443, 205-212.

Schulp, C.J.E., Burkhard, B., Maes, J., Van Vliet, J., Verburg, P.H., 2014. Uncertainties in Ecosystem Service Maps: A Comparison on the European Scale. PLoS ONE 9, e109643.

Seppelt, R., Dormann, C.F., Eppink, F.V., Lautenbach, S., Schmidt, S., 2011. A quantitative review of ecosystem service studies: approaches, shortcomings and the road ahead. Journal of Applied Ecology 48, 630-636.

Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M., Mandle, L., Hamel, P., Vogl, A.L., Rogers, L., Bierbower, W., 2015. InVEST 3.2.0 User's Guide. The Natural Capital Project, Stanford.

Sharp, R., Tallis, H.T., Ricketts, T., Guerry, A.D., Wood, S.A., Chaplin-Kramer, R., Nelson, E., Ennaanay, D., Wolny, S., Olwero, N., Vigerstol, K., Pennington, D., Mendoza, G., Aukema, J., Foster, J., Forrest, J., Cameron, D., Arkema, K., Lonsdorf, E., Kennedy, C., Verutes, G., Kim, C.K., Guannel, G., Papenfus, M., Toft, J., Marsik, M., Bernhardt, J., Griffin, R., Glowinski, K., Chaumont, N., Perelman, A., Lacayo, M., Mandle, L., Hamel, P., Vogl, A.L., Rogers, L., Bierbower, W., 2016. InVEST 3.3.0 User's Guide. The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy and World Wildlife Fund, Stanford.

Sharps, K., Masante, D., Thomas, A., Jackson, B., Redhead, J., May, L., Prosser, H., Cosby, B., Emmett, B., Jones, L., 2017. Comparing strengths and weaknesses of three ecosystem services modelling tools in a diverse UK river catchment. Science of the total environment 584–585, 118-130.

Shepherd, B., Harper, D., Millington, A., 1999. Modelling catchment-scale nutrient transport to watercourses in the U.K. Hydrobiologia 395, 227-238.

Shi, J., Davis, R., Densham, J., 2006. Better Land for Better Water: Modelling land-use change to improve water quality in England. RSPB/ WWF/ Water UK.

Smith, R.V., Jordan, C., Annett, J.A., 2005. A phosphorus budget for Northern Ireland: inputs to inland and coastal waters. Journal of Hydrology 304, 193-202.

Tague, C.L., Band, L.E., 2004. RHESSys: Regional Hydro-Ecologic Simulation System—An Object-Oriented Approach to Spatially Distributed Modeling of Carbon, Water, and Nutrient Cycling. Earth Interactions 8, 1-42.

Tallis, H., Kareiva, P., Marvier, M., Chang, A., 2008. An ecosystem services framework to support both practical conservation and economic development. Proceedings of the National Academy of Sciences 105, 9457-9464.

Tanguy, M., Dixon, H., Prosdocimi, I., Morris, D., Keller, V., 2014. Gridded estimates of daily and monthly areal rainfall for the United Kingdom (1890–2012)[CEH-GEAR]. NERC Environmental Information Data Centre.

Terrado, M., Acuña, V., Ennaanay, D., Tallis, H., Sabater, S., 2014. Impact of climate extremes on hydrological ecosystem services in a heavily humanized Mediterranean basin. Ecological Indicators 37, 199-209.

Vigerstol, K.L., Aukema, J.E., 2011. A comparison of tools for modeling freshwater ecosystem services. Journal of Environmental Management 92, 2403-2409.

Villa, F., Voigt, B., Erickson, J.D., 2014. New perspectives in ecosystem services science as instruments to understand environmental securities. Philosophical Transactions of the Royal Society of London B: Biological Sciences 369, 20120286.

Wilby, R.L., Orr, H.G., Hedger, M., Forrow, D., Blackmore, M., 2006. Risks posed by climate change to the delivery of Water Framework Directive objectives in the UK. Environment International 32, 1043-1055.

Willcock, S., Hooftman, D., Sitas, N., O'Farrell, P., Hudson, M.D., Reyers, B., Eigenbrod, F., Bullock, J.M., 2016. Do ecosystem service maps and models meet stakeholders' needs? A preliminary survey across sub-Saharan Africa. Ecosystem Services 18, 110-117.

Williams, R.J., Keller, V.D.J., Johnson, A.C., Young, A.R., Holmes, M.G.R., Wells, C., Gross-Sorokin, M., Benstead, R., 2009. A national risk assessment for intersex in fish arising from steroid estrogens. Environmental Toxicology and Chemistry 28, 220-230.

Withers, P.J., Jordan, P., May, L., Jarvie, H.P., Deal, N.E., 2014. Do septic tank systems pose a hidden threat to water quality? Frontiers in Ecology and the Environment 12, 123-130.

Withers, P.J.A., Lord, E.I., 2002. Agricultural nutrient inputs to rivers and groundwaters in the UK: policy, environmental management and research needs. Science of the total environment 282, 9-24.