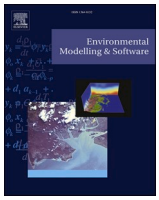
Environmental Modelling and Software 193 (2025) 106657

Contents lists available at ScienceDirect

Environmental Modelling and Software

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



Validation of the InVEST nutrient retention model across Europe with attribution of model errors

Danny A.P. Hooftman [a,b], Guy Ziv [c], Paul M. Evans [b], James M. Bullock [b,*]

[a] Lactuca: Environmental Data Analyses and Modelling, Diemen, the Netherlands
[b] UK Centre for Ecology & Hydrology, Wallingford, UK
[c] University of Leeds, UK

ARTICLE INFO

Keywords:
Agriculture
Ecosystem services
Eutrophication
Fertiliser
Grassland
Land cover
Nutrient retention

ABSTRACT

Intensive fertilisation of farmland leads to substantial nutrient escape into the environment, polluting land, water, and the atmosphere. We used the InVEST NDR model to investigate nitrogen (N) and phosphorus (P) run-off and retention across the European continent at 25 × 25 m resolution, and validated outputs against empirical measurements at 2251 river locations. Mean nutrient retention across Europe was estimated as 93 % for N and 92 % for P, through accumulation by standing vegetation and the soil. Modelled nutrient export to streams matched well to empirical measurements. Model-based uncertainties were related to seasonality, the balance between surface and sub-surface flows, and extremes in slope and rainfall. Uncertainties related to empirical data suggested enhancements to monitoring programmes that would improve nutrient export and erosion modelling, which included higher resolution fertiliser and manure data, differentiation of grassland types, including stocking density categories, and in-river nutrient measurements at low flow.

1. Introduction

Nutrients such as nitrogen (N) and phosphorus (P) are essential for crop growth, but run-off of excess N and P from fertiliser use causes land, water and air pollution (de Vries, 2021; Bullock et al., 2024). Although nutrient use efficiency of crops has improved since the 1990s (van Grinsven et al., 2014), it is far from sufficient to avoid substantial excesses of nutrients in the wider landscape. Nutrient run-off accelerates climate change (UNEP, 2019), affects human health (Henschel and Chan, 2013) and impoverishes ecosystem functioning and biodiversity (Good and Beatty, 2011; Midolo et al., 2019), as well as lowering future productivity through soil acidification (Tian and Niu, 2015).

Much environmental policy in Europe is made at the continental scale, yet implementation needs to take account of high variation in amounts of nutrient addition both within and across countries (Ludemann et al., 2022), with highly localised patterns of nutrient run-off (EEA, 2020; de Vries et al., 2022) influenced by local environmental conditions. Furthermore, the locations suffering the most adverse effects of nutrient run-off from agriculture are not necessarily the same as where the largest excesses are recorded. This is the case, for example, when nutrient run-off from land goes into waterbodies, either directly across the surface or indirectly via groundwater flows and atmospheric transport. Typical impacts of water-transported nutrients are algal blooms in coastal waters linked to upstream sources (Gilbert et al., 2018) and eutrophication of protected natural sites through atmospheric depositions (Schoukens, 2017). A critical need for policy-makers is to identify key source areas to be targeted by policy and planning (de Groot et al., 2012), rather than focusing on locations where problems result. Spatially-explicit modelling can provide credible information and maps of the spatial distribution of natural resources and environmental challenges, particularly since empirical observations usually provide only partial spatial coverage or at spatial or temporal resolutions that are less relevant to decision making (Maes et al., 2012; Hooftman et al., 2022). While measurements of nutrient excesses are hard to quantify spatially, modelling can help identify likely sources and key environmental drivers of excess loads including variation in nutrient retention ecosystem services. Furthermore, mapped modelling efforts can be used for scenario analyses (Martínez-López et al., 2019; Guaita-García et al., 2020); for instance, by assessing future impacts of agri-environment schemes on nutrient run-off in agricultural landscapes.

There are many local field scale models for nutrient use and excess (Vereecken et al., 2016), but these are not applicable at large scales, such as Europe. Ecosystem service modelling provides promising

* Corresponding author.
E-mail address: jmbul@ceh.ac.uk (J.M. Bullock).

https://doi.org/10.1016/j.envsoft.2025.106657
Received 23 April 2025; Received in revised form 13 August 2025; Accepted 19 August 2025
Available online 20 August 2025
1364-8152/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

large-scale approaches (Pascual et al., 2017; Willcock et al., 2023), where the focus shifts to the role of the natural environment in mitigating negative effects. In the case of excess nutrient run-off, the most direct contribution of the environment is the retention capacity of the vegetation and soil combined. There are several models able to simulate nutrient run-off at policy relevant scales (Redhead et al., 2018). The InVEST Nutrient Delivery Ratio (NDR) model is one of the most frequently used, validated and discussed in terms of its qualities and drawbacks (Redhead et al., 2018; Mandle and Batista, 2024).

An important aspect of modelling over large-scales is that models are variable in the quality of their predictions depending on the underlying equations and parameterisations (Willcock et al., 2019): models are interpretations of reality based on a limited set of derivable parameters. For policymakers, not knowing the model accuracy for any region of interest has a number of risks including ineffective decision-making and a reluctance to use modelled information at all (Willcock et al., 2016; Dubois et al., 2020). Ideally, every modelling effort should be accompanied by an estimate of certainty, for example by validation against empirical data (Pereira et al., 2025). Information on uncertainty helps policy- and decision-makers to develop realistic expectations in relation to their own attitudes to risk (Schuwirth et al., 2019). In such, the focus is not necessarily on an exact matching to the absolute truth, as empirical data used for validation can itself vary substantially in quality, but a reasonable and consistent estimate of uncertainty (Willcock et al., 2020).

A constructive use of the validation process is to attribute the main causes and patterns of differences between modelled and empirical data. This can lead to better understanding of where in a modelled area one should place less or more trust in model outcomes (Prestele et al., 2016). Furthermore, correlative linking of discrepancies between a model and empirical data to environmental variables can highlight key weaknesses in the model and/or potential quality issues in the validation data (Willcock et al., 2019). Both would be valuable for guiding improvements to the model or to empirical monitoring, respectively.

In this paper, we parameterise the InVEST NDR model (Scharp et al., 2018) to estimate run-off of N and P from agricultural fields and the influence of land use and management on eventual loading of waterbodies across Europe. Since within stream measurements are influenced by nutrient additions from multiple sources, agricultural and non-agricultural, we adapted the NDR model for atmospheric depositions, as well as nutrients from human wastewater. Our modelling aims to support European policy and to inform on drivers of geographical variation in accuracy. We especially focus on effective and informative approaches to generating the input parameters across this large region. We validate our models against multiple measurements of N and P in European rivers at 2251 locations (EEA, 2022). The location-specific differences between model and empirical data –the errors– are regressed against environmental drivers and model inputs that could be the source of uncertainty. We test whether substantial parts of the patterns of errors could be explained by drivers divided in six categories: (1) Validation inputs; (2) Model inputs; (3) Climate drivers; (4) Water chemistry; (5) River characteristics; and (6) Landscape characteristics. In total we tested for 28 drivers in this paper. This analysis highlights linkages between location-specific accuracy and local drivers, allowing greater confidence in using the modelled estimates.

2. Methods

Nutrient flow to streams – *i.e.*, flowing waterbodies, including rivers and streams - was modelled using the NDR module of InVEST (Scharp et al., 2018). We modelled both Nitrogen (N) and Phosphorus (P) export. A concise description of the model and the inputs employed is provided in SI-1. Fig. 1 provides an overview of the whole modelling and validation process. Our analysis focused mainly on nutrient export from agricultural areas, including differences among types of crops, but also included excess N from atmospheric deposition and human waste. The modelled year was 2018, as we utilised the crop map of d'Andrimont (2021) describing the 2018 situation. Therefore, for all sources, we used data for 2018 as far as feasible. The modelling was conducted at the resolution of the utilised Digital Elevation Model DEM (EEA, 2016): 25 × 25 m, with a GRS-1980-IUGG-1980 Lambert Azimuthal Equal Area projection (EPSG 9820). The DEM was used as the baseline resolution since it is the key factor that drives run-off movement patterns through the landscape (SI-1 & SI-4) while being least likely to change over time. By contrast, the 10-m parcel borders from d'Andrimont (2021) are likely more variable among years. All inputs were EPSG 9820 projected and were resampled to the exact gridcell size and extent as the combined Land Cover map employed (see below). Grid values across resolutions were recalculated using the ArcPro Cubic-resampling procedure.

All non-InVEST calculations were performed using ArcPro v3.2.2. (©ESRI), using python v3.9.16 for loops and command combinations; 'extractions' below refer to *Zonal*-tool usage. Matlab 9.14.0 (©The MathWorks) was employed to combine tables and for validation statistics. Jupyter Notebook and Matlab codes can be accessed via github.com/dhooftman72, which also provides all maps shown as figures in the main text and supplementary information in 600 dpi resolution.

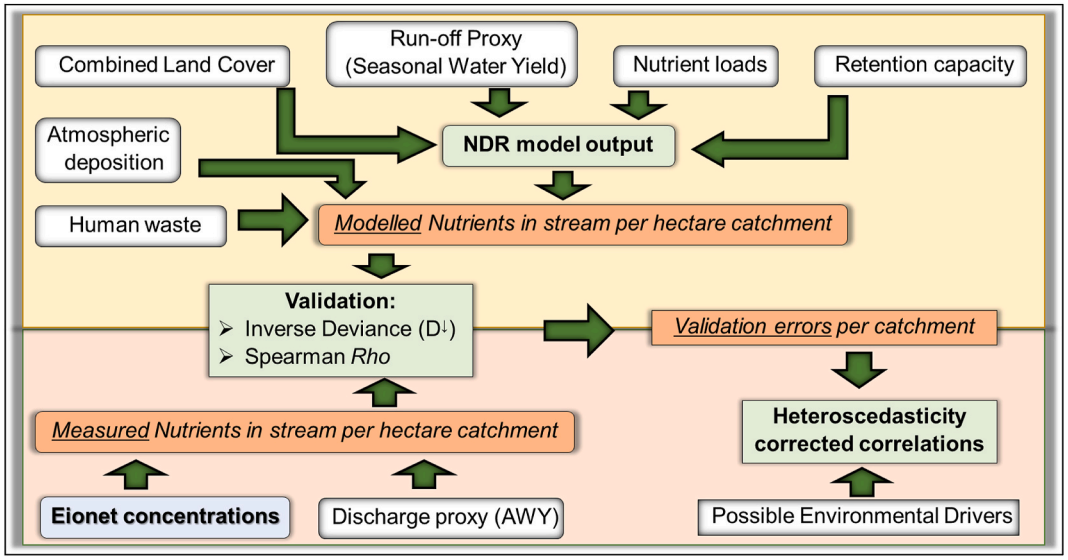


Fig. 1. Overview of the modelling, validation and correlation analyses.

2.1. Modelled area

The area modelled was the whole European Union (EU), along with the associated countries of the United Kingdom, Switzerland, Norway and Iceland, as well as the six countries formerly comprising Yugoslavia. Country outlines followed International Boundaries Level 0 (FAO, 2015). To ensure complete modelling of cross-border catchments, we also included relevant parts of Russia (including Kaliningrad), Belarus and Ukraine.

2.2. The vegetation classes map

Our land cover base was the EU Crop Map 2018 (d'Andrimont, 2021), which distinguishes among 18 crop types and one grassland type, and was *majority-aggregated* from 10 m to 25 m resolution gridcells. The EU Crop Map covers only agricultural areas. To fill the gaps for non-agricultural vegetation, woodlands and urban areas, we employed the ESA WorldCover 2020 map (ESA, 2020). European non-EU countries not covered by the EU crop map were filled with the same WorldCover map, with a single cropland category.

The EU Crop Map 2018 does not differentiate between 'improved grassland' and 'non-improved', *i.e.*, grasslands to which synthetic fertilisers are not applied. To differentiate these grasslands, we used EEA (2020), which describes a nutrient flow model with an emphasis on improved grasslands. Gridcells were set to non-improved when they were 'no-data' in EEA (2020, *i.e.*, no agricultural relevance) combined with a grassland definition in WorldCover (ESA, 2020). We refer to this combination of EU Crop Map and WorldCover maps as the Land Cover Map. This map is depicted in SI-2.

Input loads are available at the country level (FAO, 2015) and were reflected in our modelling by subdividing the Land Cover Map per country. Moreover, since the proportion of subsurface flow depends on slope, five slope classes were generated from the European Digital Elevation Model (DEM; EEA, 2016; SI-3). In alignment with the InVEST Seasonal Water Yield (SWY) model (see below; SI-4), four hydrological soil groups were included (Ross et al., 2018). The combination of these groupings generated a Vegetation Class map and table of 12,112 units; the table is given in SI-5 with associated parameter values.

2.3. Nutrient run-off proxy via Seasonal Water Yield

The NDR requires a Nutrient Run-off proxy representing the capacity to transport nutrients downslope to the streams. In its most simple form this could be set as annual precipitation (*e.g.*, Redhead et al., 2018). We applied a more nuanced option, employing the InVEST Seasonal Water Yield model (SWY; SI-4). This allowed us to calculate the proportion of surface ('quickflow') compared to subsurface flow ('baseflow'), which the NDR model requires. Sharp et al. (2020) describe the full set of SWY equations, and Bagstad et al. (2018, 2020) describe its use in detail. Input variables for the SWY model are listed in SI-4. The sum of quickflow and baseflow combined was set as the total amount of water reaching the stream as the Run-off input for the NDR model. The median proportion of all flow that was baseflow (SI-5) was extracted as the mean over all gridcells for each of the 12,112 vegetation class units.

2.3.1. Agricultural nitrogen and phosphorus loads per vegetation class

The nutrient amount that reaches the stream depends on quantities of synthetic fertiliser and manure applied and excreted on agricultural fields. Atmospheric deposition and human waste were added to this, as described in 2.3.2. We considered the excess nutrients –nutrients not absorbed by the crops– as potential run-off, termed 'nutrient load'. All input data used to calculate loads were per country per crop type.

Country-averaged synthetic fertiliser applications per crop type for N and P were derived for 2018 from Ludemann et al. (2022). Limited bespoke adaptations included tonnage weighted averaging among different maize types (food, feed, biomass) and vegetables which we set as equivalent to 'other non-permanent industrial crops'. Manure application followed the EuropeAgriDB v1.0 database for N (Einarsson et al., 2021). The categories 'Applied to cropland' and 'Excreted grazing on cropland' were equated to cropland and added to each cropland type equally. The category 'Excreted grazing on permanent grassland' was similarly added to non-improved grasslands, whereas the categories 'Applied to permanent grassland' and 'Excreted while grazing' were added to improved grasslands. P from manure was estimated with an N:P ratio of 5.3:1 (Oenema et al., 2021).

Excess nutrients, the 'nutrient loads' for N and P separately, were calculated as:

$$Load = (1 - Nutrient\ Use\ Efficiency) \times (Fertililiser + Manure)$$

Here, fertiliser and manure are the applied gross amounts of nutrients to the fields in tons ha^{-1}. The Nutrient Use Efficiencies, as proportions, were derived from EEA (2020) for N and P separately. We extracted values as medians for the Land Cover Map classes combined with four hydrologic Soil Groups (Ross et al., 2018). Further subdivision of these resulting 140 categories into all 12,112 vegetation classes was not possible because of the relatively coarse 1 km resolution of the EEA (2020) dataset. We therefore extrapolated values as equal within the further subdivision (SI-5).

Agricultural loads are shown as maps in Figure SI-2-1. Data above included the EU member states and the UK. For other associated non-EU countries, no among-crop type distinction is available, and the load information described above is unavailable. To allow their inclusion in the modelling and the sources of nutrients they provide, loads were based on values from similar countries (listed in SI-3).

2.3.2. Atmospheric deposition and human waste

We included atmospheric deposition of N as important source of excess N (EMEP, 2022; depicted in SI-2). Atmospheric deposition was run as a separate NDR model, the output of which was added to the fertiliser model described above. As load inputs, we used combined dry and wet deposition of NO_x and NH_x from EMEP (2022). Since deposition patterns are at sub-country level, we divided the EMEP (2022) dataset into 20 classes (SI-3). These classes were combined with slope classes (SI-3) and a simplified version of the Land Cover Map in six classes (SI-3). For each of the 595 resulting vegetation classes the median deposition was extracted. The Nutrient Use Efficiencies were extracted as for the fertiliser model. This table can be found in SI-5. For all other parameters, this deposition model followed the fertiliser model. Atmospheric loads are depicted as map in Figure SI-2-3.

A further source of nutrients is that of human waste via sewage. We used a mean annual per capita export of P and N in untreated sewage of 0.52 kg P and 4.5 kg N (Naden et al., 2016; Redhead et al., 2018), with a waste plant efficiency of 60 % and 90 % for N and P respectively (Li et al., 2017). Unconstrained population counts per hectare from WorldPop for 2018 (Lloyd et al., 2019) were multiplied with these constants. In the validation analysis, the per catchment extractions of human waste were added to the extracted outputs of the fertiliser and atmospheric deposition models. As wastewater excretion would be from point sources – unavailable at this scale – wastewater is only included in the validation analysis at the catchment scale and not per gridcell, assuming at least one outlet per catchment.

2.4. Retention capacities

The NDR model is sensitive to the 'retention capacity' of the vegetation. However, very few useable data exist to estimate this value. Studies such as Redhead et al. (2018), Zawadzka et al. (2019) and Lavorel et al. (2022) were restricted in their sources and so in variation among vegetation types. Here, we explored a spatial approach, making the amount of retention related to the Normalised Difference Vegetation Index (NDVI), which is a measure of the density of the vegetation and thus its ability to trap and use nutrients. We used 2018 16-day NDVI

values from the Modis Terra satellite data averaged per year. We developed a bespoke power calculation for the 41 vegetation classes of the Land Cover map (SI-1) following:

$$Retention\ \% = \alpha \times (NDVI - \beta)^2$$

Subsequently, α and β were fitted such that the values for woodland, bare areas and cropland would match Redhead et al. (2018) and Zawadzka et al. (2019). We elaborate more on this method in SI-6.

Based on the load and the NDR outputs, we calculated the proportion retention per gridcell *x*, which includes nutrients accumulated and not reaching the stream, as $\left[\frac{(Load_x - Export\ to\ Streams_x)}{Load_x}\right]$. For N, both loads and export to streams are a summation of modelled outputs from manure & fertiliser and atmospheric depositions (see below).

2.5. Validation

Modelled outputs were validated against measurements in rivers from the European Environment Information and Observation Network (Eionet; Wise v6, EEA, 2022). Measurement locations were selected from the 'aggregated' data, being those locations which had more than 10 database entries of N and more than 15 for P since 1995. N was a combination of NO_x and NH_x, with their atomic weights used in recalculations to represent N alone. These concentrations were in mg per litre. To translate this to the modelled tonnages required an annual river discharge estimate to calculate an annual quantity of nutrients passing through the given points in rivers as *[median concentration x annual discharge in litres]*. As too few discharge measurement points within a short radius of the selected Eionet points were identified, we generated an InVEST Annual Water Yield model (AWY; Sharp et al., 2020), which is concisely explained in SI-7. The AWY is mathematically dissimilar to the SWY used for the NDR model as it is not based on a stream network and multiple inputs are different (SI-7). Having an annual discharge estimate implicitly assumes that the nutrient measurements are representative of the annual quantity of nutrients. We are aware that this assumption is a source of unavoidable uncertainty.

Following Willcock et al. (2023), we generated bespoke catchments with each measurement location as a catchment outlet, using the 25m EU-DEM v1.1. (EEA, 2016). We allowed for a maximum deviation distance of 100m from the EU-Hydro River Network Database (Copernicus, 2020). Annual discharge was extracted as the sum of all within catchment water yields per cell. The total number of selected validation points was 2251.

We compared the total modelled amount of N export *[fertiliser + manure + atmospheric + human waste]* and P export *[fertiliser + manure + human waste]* within catchments to the measured amount of nutrient estimated to pass through the catchment outlet. Model accuracy was assessed using the Inverse of Deviance ($D^{\downarrow}$; Willcock et al., 2019); the inverse of the mean absolute individual deviance of each data point from the 1-1 line of equal relative value. Inputs were normalised with a double-sided 2.5 % winsorising protocol, to reduce the effect of outliers, following Hooftman et al. (2022) and Willcock et al. (2019, 2023). A $D^{\downarrow}$ above 0.7 is considered to show good fit to validation data (Willcock et al., 2019). Moreover, we conducted rank-order regression using Spearman's *Rho* to test if modelled data followed the same order as the validation data (*corr*-tool with Spearman link). To avoid spurious significance by using large numbers, validation analyses were jack-knifed for 225 datapoints each (10 % of datapoints), taking the median among 10,000 runs.

2.6. Per datapoint errors

Per datapoint errors were used to attribute drivers of error and were calculated as the signed deviance = *[modelled value - validation]* of each modelled datapoint for which there was validation data, where both modelled and validation data were normalised (see above). We called this the *error*, with the error sign separating over- and underestimation. To estimate the influence of environmental drivers, errors were subjected to linear regressions using a number of potential explanatory factors (Table 1). These factors are listed in SI-8 with their data sources and explanations where applicable. These drivers relate to six categories of explanatory factors: (1) Validation inputs; (2) Model inputs; (3) Climate drivers; (4) Water chemistry; (5) River characteristics; and (6) Landscape characteristics.

Values for all drivers were calculated as median over full validation catchments. Exceptions to this were the Eionet measurements (N and P measurements and chemistry) which were calculated as the median of measurements at the river validation location. Prior to regression analyses, these drivers were normalised with a double-sided 2.5 % winsorising protocol (Hooftman et al., 2022), while the errors were normalised with a single-sided 2.5 % winsorising protocol to maintain the signed expression. In addition, each combination of driver and error

Table 1
Regression coefficients for the effects of different drivers on errors for model differences from validation data. Original driver units are provided, but note data were normalised before analysis. See SI-8 for driver descriptions. For significant effects, the number in parentheses is the value at which there is a cross-over between over- and underestimation (for example: for Median Total mg/l, Nitrogen: predicted cross-over from over-to underestimation is at 29 % of the data-range). A Hochberg correction is employed to correct for multiple tests following Willcock et al. (2023).

	Nitrogen Error	Phosphorus Error
Validation set at validation location		
Median Total mg/l (=validation)	−0.84[a] (0.29)	−1.10[a] (0.11)
Coefficient of Variation of measurements	−0.26[ns]	0.01 [ns]
Discharge AWY model (m^3 ha^{-1})	0.18[ns]	0.39[a] (0.26)
Model Inputs		
Flow from SWY model ha^{-1} (m^3 ha^{-1})	0.33[c] (0.01)	0.40[a] (0.16)
Load from soil fertilisation & manure (t ha^{-1})	−0.03[ns]	0.37[b] (0.11)
Atmospheric deposition N load (t ha^{-1})	−0.39[c] (0.43)	–
Waste water load = Human pop size (t ha^{-1})	−0.16[ns]	−0.26[ns]
Proportion sub-surface flow	0.36[a] (0.44)	0.18[b] (0.55)
Climate		
Rain erosivity index (MJ mm ha^{-1} h^{-1} yr^{-1})	−0.20[ns]	0.36[a] (0.29)
Median Daily temperature (°C)	−0.59[a] (0.59)	−0.03[ns]
Median # Ice days $year^{-1}$	0.58[a] (0.14)	0.59[a] (0.18)
Chemistry at validation location		
pH of the water	−0.32[c] (0.80)	−0.02[ns]
Concentration Suspended Carbon (mg/l)	0.45[a] (0.21)	0.19[ns]
Concentration dissolved O_2 (mg/l)	0.05[ns]	−0.37[b] (0.38)
River characteristics		
Stream curviness (sinuosity: actual length/ direct line length)	0.16[ns]	0.09[ns]
Stream gradient (°, proxy for stream speed)	−0.05[ns]	0.41[a] (0.18)
Median distance to Stream (m $catchment^{-1}$)	0.01[ns]	0.05[ns]
Catchment size (hectares)	−0.01[ns]	0.10[ns]
Terrain Ruggedness (STD of ° slopes)	−0.23[ns]	0.31[a] (0.33)
Landscape characteristics (% of the landscape per catchment)		
Woodland	−0.11[ns]	0.25[ns]
Cropland	−0.04[ns]	−0.13[ns]
Improved grassland	−0.14[ns]	−0.22[c] (0.21)
Not-improved grassland	0.27[c] (0)	0.56[a] (0.18)
Urban/artificial surfaces	−0.06[ns]	0.11[ns]
High fertilised crops (SI-8 for list)	−0.06[ns]	−0.13[ns]
Low fertilised crops (SI-8 for list)	−0.10[ns]	0.01[ns]
Density of Cattle (stock)	−0.34[c] (0.33)	−0.17[ns]
Density of Sheep (stock)	−0.29[ns]	0.02[ns]

ns: non-significant, p-levels after correction.
[a] $P < 0.001$.
[b] $P < 0.01$.
[c] $P < 0.05$.

was *a priori* subjected to a linear heteroscedasticity correction, based on Addy et al. (2022), which is explained in SI-9. Linear regressions for each individual driver were carried out using the *fitnlm*-tool employing a weighted standard linear function [$Y \sim b_1 + b_2X$], with the proportion overlap among catchments as datapoint weights (SI-10). To avoid spurious significance by using large numbers, the regressions were jack-knifed for 225 datapoints each, taking the median among 10,000 runs. We employed a Hochberg's step-up correction (Huang and Hsu, 2007) on the resulting average p values to account for multiple tests.

To determine whether each driver primarily causes over- or under-estimation, we calculated the regression cross-over point. This is the driver value at which the regression line intersects the "is-equal" line. As drivers were normalised this cross-over value corresponds to the proportion of the range that is over- or underestimated. Based on b_1 and b_2 this predicted cross-over point (Y_0) from over-to underestimation or *vice versa* was calculated as $Y_0 = \frac{-b_1}{b_2}$. Furthermore, a correlation matrix was generated among all drivers using the *corr*-tool, jack-knifed as above, presented in SI-11.

3. Results

3.1. Nutrient losses

Fig. 2 presents a Europe-wide 25 × 25 m estimates of nutrient losses to streams (*i.e.,* flowing waterbodies). Nitrogen (N) losses to streams result from added fertiliser, manure, and atmospheric deposition of NO_x and NH_x (Fig. 2a). Phosphorus (P) losses are from fertiliser and manure (Fig. 2b), the former generally added as P_2O_5.

Absolute nutrient losses to streams are large from grasslands with high input loads (Figure SI-2-1), such as in The Netherlands, Ireland, Western Denmark and parts of France and Germany (Fig. 2). Combined N loads are further elevated by high atmospheric nutrient deposition, especially in The Netherlands, Belgium, North-Western and Southern Germany, and Danish Jutland (Figure SI-2-2). For the same reason cities have high exports: urban landscapes have low retention potential and receive substantial atmospheric nutrient deposition in especially North-Western Europe. Fig. 2 excludes wastewater nutrients (which is taken into account in the validation analysis), which would accentuate the role of cities even more. Atmospheric N deposition contributes 22 % of the total load across Europe; 58 % of N originates from agricultural N and 20 % from filtered human waste. However, there is large spatial variation, and for the majority (63 %) of the modelled area, atmospheric N deposition is the dominant source (Figure SI-2-3).

Visually, large nutrient export values are seen in grassland-dominated areas in which substantial variation in the terrain are combined with high rainfall, such as in the Northern UK, North-Western Spain and in Switzerland (Fig. 2). This is substantiated by high positive correlations between nutrient export per hectare with the per catchment proportion of non-improved grassland, density of cattle, and water run-off (correlation coefficients of 0.41, 0.30 and 0.56 respectively for N, SI-11). Grassland effects are more pronounced for N (Fig. 2a) than for P (Fig. 2b). By contrast, in most of Southern and most areas of Eastern and further Western Europe differences between crop types seem more pronounced, being less terrain and rainfall dominated. Over substantial parts of Southern Europe, the low to intermediate fertiliser levels on permanent crops such as olive trees and fruit trees seem to lead to relatively low nutrient export; comparing Fig. 2 with the Land cover Map shown in Figure SI-2-4. Low nutrient losses are found in regions without wide-spread intensive agriculture and low atmospheric deposition (Figure SI-2-2) such as in heavily forested parts of Scandinavia, the Baltic states and Southern European regions; as well in other areas with high forest or alpine vegetation. This is further substantiated by the per catchment proportion of forest being negatively correlated with nutrient export per hectare (correlation coefficients of −0.39 and −0.27 respectively for N and P, SI-11).

3.2. Retention

The proportion of nutrients retained is similar for N and P (Fig. 3). On average, retention across the whole modelled area is 93 % for N (STD 17 %) and 92 % for P (17 %), *i.e.,* 7 % and 8 % of excess nutrients reach the streams respectively. For N this is retention both by the vegetation and by the soil filtering the nutrients from the subsurface flow. The retention capacity seems lower where there is a combination of high terrain variability with high rainfall, causing the higher export to streams described above. In those cases, the proportion of surface run-off is high compared to baseflow run-off – which can allow filtering out of nutrients by the soil – in areas with lower variability of the terrain. Furthermore, the modelled retention capacity in areas with low nutrient inputs, as in

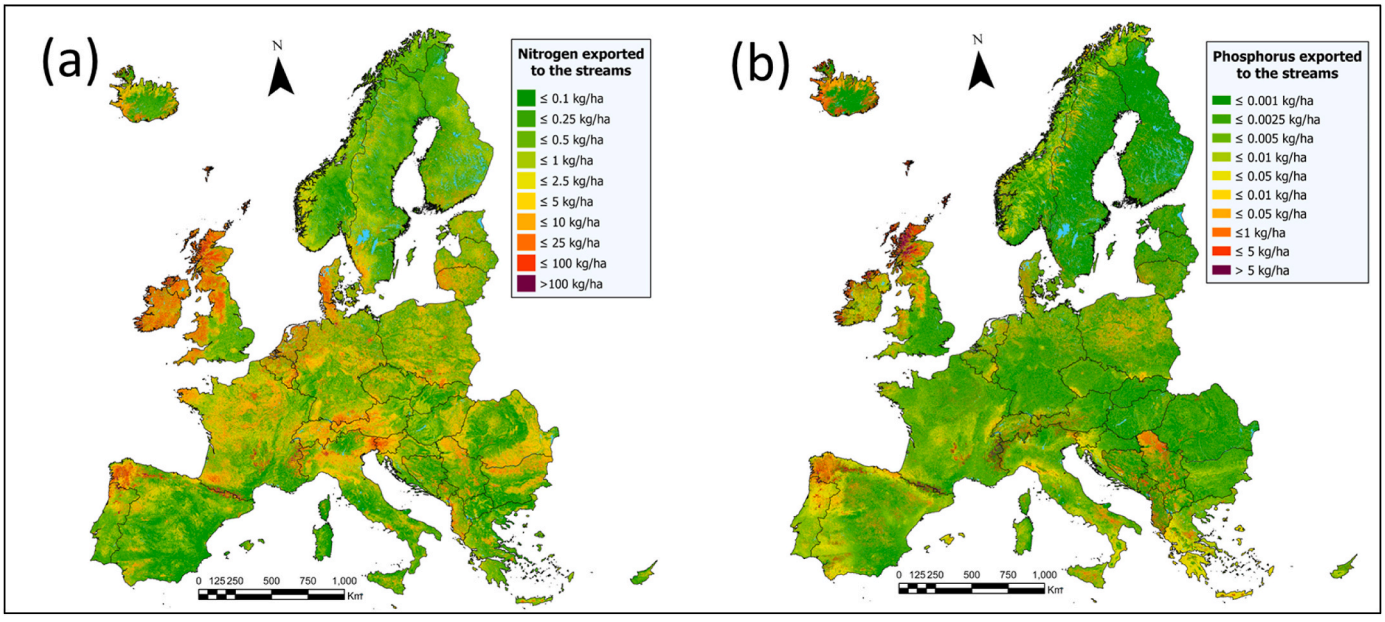


Fig. 2. European wide estimates of nutrient losses to the stream from fertiliser and manure additions combined with (for Nitrogen only) atmospheric deposition generated with the InVEST Nutrient Delivery Ratio model for (a) Nitrogen and (b) Phosphorus.

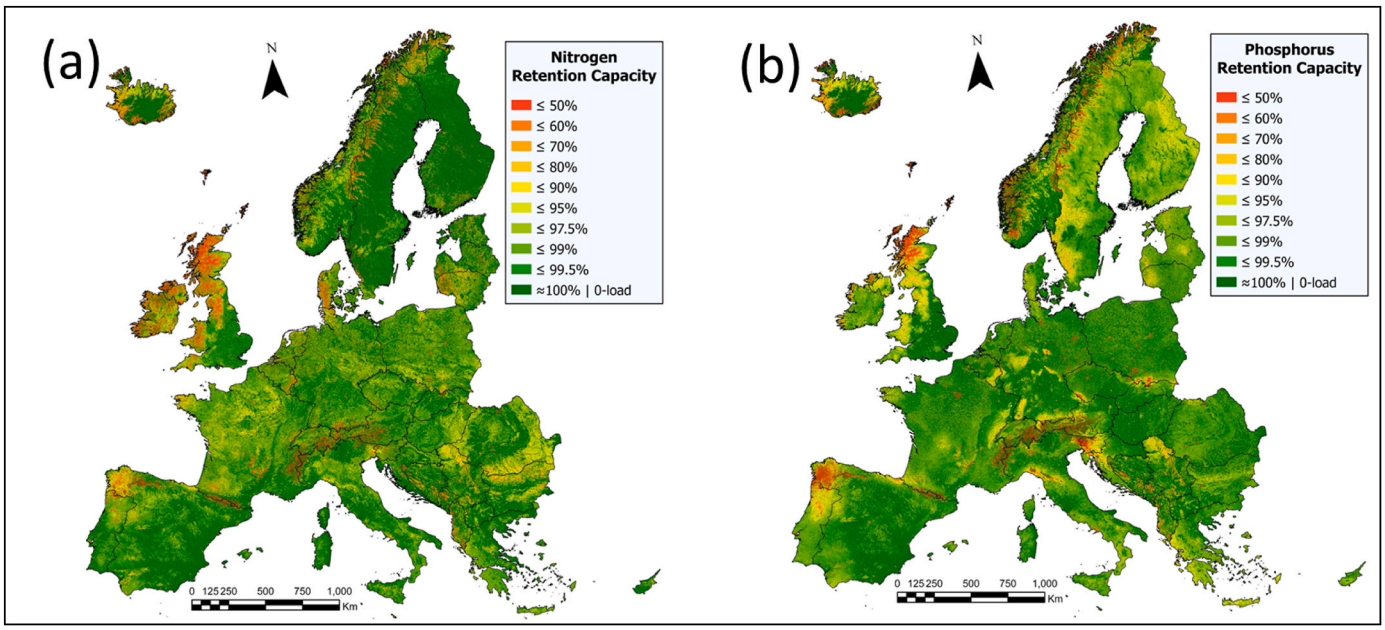


Fig. 3. European wide estimate of the retention capacity of the vegetation as proportion of retained loads from fertiliser & manure additions and atmospheric deposition (Nitrogen only) for (a) Nitrogen and (b) Phosphorus.

Scandinavia excluding Denmark, is relatively low. This is, however, a low retention rate of low amounts of nutrients.

3.3. Validation

The NDR models were validated against 2251 nutrient concentration measurements in streams, selected from the Eionet dataset (Fig. 4). In general, the modelled medians for N and P among catchments are close to those of the validation set, 3.62 *vs.* 2.03 kg ha^{-1} catchment for N and 0.07 *vs.* 0.11 kg ha^{-1} for P. However, the 5 %–95 % range of model predictions of N is larger than that of the validation dataset while model predictions are missing very low values (range 0.38–26.0 *vs.* validation 0.03–12.5), with 76 % of the data points being overestimates. The missing low values show up as large overestimates in Fig. 5. By contrast, P is in general underestimated by the model (66 % of points are underestimated) including a series of substantial underestimations (Fig. 5; range 0.01–0.64 *vs.* validation 0.01–0.97) caused by a set of high validation outliers, distributed across Europe without apparent pattern (Fig. 4).

Fig. 5 depicts the validation as kg nutrient per hectare catchment. The validation of N is very good (Fig. 5a), with a high Inverse of Deviance ($D^{\downarrow} = 0.82$): indicating an overall close proximity to the 1:1 line. The overestimation seems consistent with a median absolute difference of 78 %, decreasing to equal values at high amounts of predicted N. The $D^{\downarrow}$ is lower in Swedish, Finnish and Norwegian catchments with low measured concentrations, which are overestimated by the model (see SI-

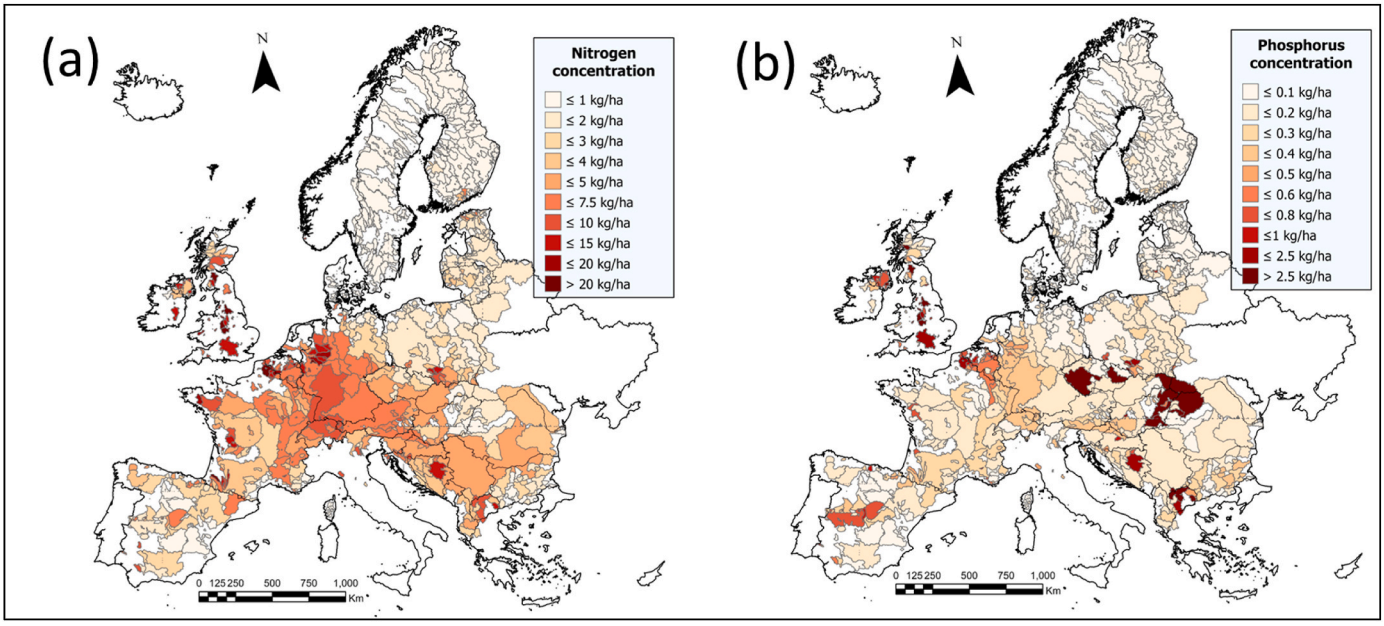


Fig. 4. Delineated catchments from selected Eionet locations (EEA, 2020) used for validation, with (a) median averaged Nitrogen measurements; (b) median averaged Phosphorus measurements. n = 2251. Catchments can overlap, which is taken into account in our analyses (SI-10).

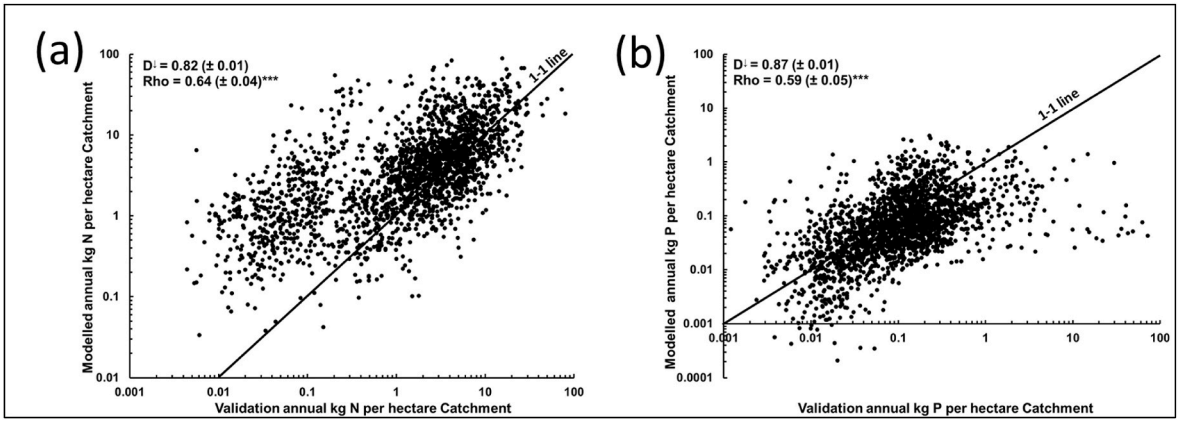


Fig. 5. Validation of the modelled annual kg per hectare catchment passing through validation locations compared to the validation data as annual kg per hectare. For (a) Nitrogen and (b) Phosphorus. n = 2251; validation statistics are added.

12). The rank correlation (*Rho*) is highly significant (0.64; $P < 0.001$), showing that the order of N export among catchments is predicted well. However, the model performs more poorly in certain countries (Table SI-12-1). As well as the overestimation of the very low Scandinavian validation values, values for Ireland and Denmark are overestimated by 8 and 10 times respectively (Table SI-12-1).

The validation for modelled P is very good. Except for a set of high validation dataset outliers (Fig. 4), the proximity to the 1-1 line is high ($D^{\downarrow} = 0.87$; Fig. 5b). Because of these outliers, there is an average underestimation by 35 %. The rank correlation (*Rho*) is also highly significant (0.59; $P < 0.001$), showing that the order of P export values among catchments is predicted well.

3.4. Drivers associated with per datapoint errors

Per datapoint errors tend to shift from over-to underestimation with increasing values of the validation data, both for N and P (Fig. 5). This is confirmed by the regressions with validation values (Table 1; Hoghberg corrected *p* values), with a transition from over-to underestimation predicted at relatively low values.

Both N and P errors are related to drivers from all six categories (Table 1), although for river and landscape characteristics, the exact driver varied. For the shared drivers among three further categories, all show positive effects – *i.e.*, underestimation becoming overestimation at larger values – and this is seen for the amount of nutrient exporting flow, the proportion of sub-surface flow (both of these are model inputs), the number of ice days (climate driver) and the proportion of non-improved grasslands (landscape characteristic). For all, transitions to overestimations were predicted at relatively low range values, except for the proportion of sub-surface flow. This suggests the error increases with increasing driver value over most of the driver range.

In addition to the general patterns described above, N errors are related to the atmospheric deposition values with higher depositions linking to underestimations (model inputs). Furthermore, the model underestimates more in warmer, higher pH and low organic matter (carbon) river conditions, and overestimates as these driver values move to the opposite values (chemistry; Table 1). Moreover, underestimations are associated with greater cattle densities, interacting with the overestimations described above that are related to increasing amounts of non-improved grasslands (landscape characteristics).

Overestimations for P are further linked to high fertiliser and manure loads (model inputs; Table 1). Moreover, overestimations are linked to high flows (model input), high erosivity (climate) as well as steep stream gradients and rugged terrain (river characteristics), with cross-overs from under-to overestimation often at the lower end of the data ranges. With respect to landscape drivers, P errors depend on an interplay between improved (underestimation at high values) and non-improved grasslands (overestimation at high values).

4. Discussion

4.1. The modelled maps

We have presented a Europe-wide parameterization of the InVEST Nutrient Delivery Ratio model (Sharp et al., 2020), quantifying water-based run-off and retention of nitrogen (N) and phosphorus (P), which matches well with measured validation data. We have expanded on the methodology and spatial extent compared to earlier studies such as Redhead et al. (2018), Zawadzka et al. (2019), and Lavorel et al. (2022). Across Europe, models such as this can be used to indicate regions of most concern, either because absolute export is high, or retention capacity is low. The total summed export to streams over the modelled area is around 2.3 Tg N per year and 0.07 Tg P, although this excludes wastewater, which mostly comes from urban areas. These total values are in line with the Eionet-trained catchment based European JRC GREEN model (Grizzetti et al., 2021), which was applied over a ca. 60 % larger area (≈2.9 Tg N and 0.12 Tg P from agriculture and atmospheric deposition; Grizzetti et al., 2021).

The model outputs matched well with within-river measurements of N and P (Eionet, EEA, 2022). In terms of both the absolute units and normalised, relative units, the model values are not dissimilar to the validation data, with a mean 18 % and 13 % normalised deviance for N and P respectively. For validation at such a scale over such a heterogenous area, this can be seen as an encouraging result, since variation can be large due to catchment-specific processes (Edwards and Withers, 2008; Redhead et al., 2018). Lower deviances might be expected when validating over a more uniform and smaller area, which would allow for better parameter training and sensitivity analyses (Willcock et al., 2019). Here, we identified model and environmental drivers that seem to influence modelling errors at the catchment scale, noting that the validation data will also be subject to error. A major finding is that nutrient pollution in the high production agricultural regions is predominantly caused by fertiliser use. However, for 63 % of the modelled area, including most forested and alpine areas, the main source of N pollution is indirect: through atmospheric deposition derived from excesses elsewhere (SI-2). Results like ours will also be useful for decision-makers when interpreting modelled data over larger scales beyond single catchments (European Commission, 2018; Grizzetti et al., 2021). Validation, as performed here, would provide an additional level of confidence in utilising model results for decision-making, as exemplified by the parameterised InVEST NDR model used in our study.

4.2. Drivers of model errors

We found evidence that most categories of drivers can be linked to the amount and direction of model error; *i.e.*, the relative deviance of modelled values from the validation data. While we are aware that more individual drivers could be identified, our main focus was on the six main categories, as these drivers could be proxies for a larger set of similar drivers. For the first category (validation values), the NDR model errors are for both nutrients related to the amount of nutrient measured from catchments. The number of points over- and underestimated differs between N and P: 74 % of catchments overestimated for N (representing 71 % of the range) whereas 66 % underestimated for P (11 % of validation range). We discuss this mismatch for P in 4.4. below. After normalisation, model outputs for both nutrients show a similar transition from relative over-to underestimation over the gradient from low to high values in the validation dataset. The apparent causes of model error at catchment level are different though between both nutrients. For N, the model overestimates low validation values considerably. By contrast for P, high values seem to be underestimated substantially. At a smaller scale, similar patterns of underestimation of both nutrients were obtained by Redhead et al. (2018) for the InVEST NDR model applied across the UK.

All five further driver categories included at least one driver with a significant effect on with model errors for N and P: model inputs; climate; water chemistry; river characteristics; landscape characteristics. Four individual drivers across three categories were linked to errors for both nutrients: the amount of nutrient exporting flow, the proportion of sub-surface flow (model inputs), the number of ice days (climate), and the proportion of non-improved grassland (landscape characteristic). The model input and climate category drivers likely highlight the limitations of the modelling in that it employs an approach that averages over the year; this issue was also identified by Redhead et al. (2018) and Grizzetti et al. (2021). In reality, nutrient run-off is not continuous and surface run-off in general only occurs in (especially heavy) rain periods. Ice days illustrate a similar issue in that flow is restricted during these periods. Therefore, actual run-off is periodic, affected by soil saturation levels that will vary the balance between sub-and above surface flow ('quickflow'). Furthermore, the retention capacity by vegetation differs among seasons (von Schiller et al., 2008). Since such processes cause substantial variation in export, these will lead to model error.

4.3. Nutrient retention in soils

We calculated a 93 % retention of N by vegetation and soil, accumulating over a mean of 920 m before reaching the stream. Such continuing and stable soil retention as assumed by the model would require indefinite local accumulation around rooting zones and retaining vegetation (Wang et al., 2018). Such accumulation does occur to some degree, causing adverse acidification effects and biodiversity losses due to eutrophication (de Vries, 2021). However, N tends to stay in the soil for only a short time (Huntingford et al., 2022). Excess soil N is converted to harmful reactive gaseous elements, such as NO_x and NH_x. This can happen before excess N reaches the stream and also is a process that takes place within streams. Across the catchments studied, atmospheric deposition of NO_x and NH_x –converted from such excesses elsewhere– forms a substantial 22 % of total modelled N passing through catchment outlets. Terrestrial soil denitrification will also convert a sizeable proportion (30–60 %) of N to atmospheric N_2 (Wang et al., 2018). Since factors affecting the N-cycle balance between harmful NO_x & NH_x and harmless N_2 are poorly understood (Wang et al., 2018; Huntingford et al., 2022), the N-cycle is yet to be incorporated in models such as the NDR. N-cycle processes could explain why chemistry indicators (pH and suspended carbon) and daily temperature regressed significantly with N model errors. However, pH is also strongly correlated with annual temperature and ice days (SI-11), which suggests a combined spatial and chemical effect. Similarly, no nutrient leakage to the lower groundwater levels is yet included in the NDR, which could lead to nutrient pollution over much longer time-frames not included in the current retention estimation. Furthermore, the proportion retention is depending on the definition of 'stream', here modelled as a flow accumulation of at least 1000 gridcells (SI-1; 62.5 ha); a threshold set after matching to the EU-Hydro River Network Database (Copernicus, 2020). A tighter definition with less accumulating area will lead to lowered distances to streams, potentially resulting in less estimated retention capacity.

The chemical cycle for P is less complex between water run-off and in the soil accumulation (Panagos et al., 2022), with P run-off being more affected by peak flows (Edwards and Withers, 2008; Redhead et al., 2018). P modelling errors also related to flow speeds, as indicated by the relationships with river slope and terrain ruggedness – which are highly intercorrelated (SI-11) – as well as the annual summed erosivity of rainfall. The positive direction of the effects (overestimations at higher values) could indicate that peak flows might be less well captured in the validation data (Defew et al., 2013) rather than simply being poorly captured by the model. This is because for the latter the converse of transport, sedimentation, would lead to overestimations at low flow, which we did not identify. At smaller scales, in a subset sufficient discharge information, these processes could be investigated further by using time specific discharges employing a Beale Ratio as done *e.g.* in Redhead et al. (2018) and Robertson and Saad (2019).

4.4. Loads and Land Cover Map

The dependence and subsequent drawbacks of InVEST and similar approaches that assign discrete values to land cover classes is a long-recognised issue (Eigenbrod et al., 2010). To reduce tbib_eigenbrod_et_al_2010his problem, we generated over 12,000 vegetation land units over the full European area, overlaying the EU Crop Map (d'Andrimont, 2021) with country borders, slope and hydrological properties. However, several drawbacks remain.

The first main issue is the fertiliser data being available only at a country resolution. Consequently, synthetic fertiliser and manure has to be modelled as added equally across regions within a country, meaning that very low and high within-country values are not replicated in the input data. The inability of the model to mirror low validation values for N might be explained by this issue; as well as the modelled high export values for the lower slopes of rugged areas. Retention is lower in the latter areas since most run-off is from quickflow, leading to high modelled export. This issue could be solved by splitting databases as those used (Einarsson et al., 2021; Ludemann et al., 2022) into regional levels. However, at the scale of Europe, this would require a large number of assumptions. Furthermore, in using these databases there seems a mismatch between low validation values and high modelled export in Ireland and Denmark, by a factor of about 10 (SI-12). The synthetic fertiliser and manure load values are similar to surrounding North-Western European countries (SI-3), as could be expected. Therefore, this suggests there may be representability problems with the validation data, perhaps due to the specific location or timing of sampling.

The second main issue with model parameterization is that the EU Crop Map (d'Andrimont, 2021) differentiates 18 crops, but contains only one grassland category, not splitting improved and non-improved grassland. We allocated non-improved grassland using ESA (2020) and EEA (2020). However, the variability in stocking rates and fertilisation levels of grasslands could not be included. This drawback is seen in the regression analysis of model errors: the greater the area of non-improved grassland the more the relative model overestimation for N and P, suggesting the load values employed were too high. The converse relationships are consequently found for the amount of cattle (N) and proportion of improved grassland (P), suggesting under-parametrisation for intensively stocked areas. For the UK, Marston et al. (2023) would be a great resource to reflect grassland type diversity. However, similar

approaches are not available for most of the European area. Therefore, if there is to be an updated version of the EU crop Map (d'Andrimont, 2021), we suggest this should have a more nuanced split into multiple grassland categories; *e.g.*, into categories reflecting different levels of agricultural improvement, and livestock densities. This would not only give a better overview of grassland agronomic usage, but it would also allow more accurate EU-wide modelling of ecosystem services such as erosion and nutrient control.

This issue relates to a more general one concerning resolution and landscape classes in the modelling. A more detailed land cover class definition and finer resolution mapping could enhance modelling precision in more heterogeneous landscapes (Grêt-Regamey et al., 2014; Lamy et al., 2016), such as in some highly transformed, fragmented regions (Qiu et al., 2021). By contrast, using a higher resolution in simpler landscapes would be unhelpful but without obvious negatives. In general, more non-crop areas in landscapes result in less erosion and so lower modelled nutrient run-off (Duarte et al., 2018). Hence for nutrient retention, the representation of fine-scale land covers, such as through agri-environmental programs, might increase the modelled capture of nutrient run-off. It is likely, therefore, that adding finer relevant detail in land use definitions would result in more retention and a lower modelled run-off in more heterogeneous regions, which would likely increase accuracy of the NDR model. This suggestion comes with the caveat that this finer resolution data need to be spatially correct, well-scaled and accurately parameterised, otherwise error is likely to increase (Pereira et al., 2025). That is, including more land cover types and using a finer resolution comes with increased data demands.

A third issue is that NDR requires data on the retention capacity of different vegetation types. We have observed that such values are generally just copied from study to study, with the original sources often being obscure. We tried to provide a more spatial approach based on NDVI (SI-6), which follows similar approaches for vegetation cover impacts on sediment retention (van der Knijff et al., 2000). However, by having to fit to the range of values used by other studies such as Redhead et al. (2018) and Zawadzka et al. (2019), the methodological advance might be limited and the resulting variation among crop types and other vegetation types is relatively low, *i.e.*, many values are similar. Presumably, among region differences will mask any clear differences among vegetation types when averaging into cross-European class values. We call for more development in vegetation retention calculation possibilities.

5. Implications

At a European level, gridcell-based maps with complete coverage, as presented in this work, are valuable for identifying regions where more detailed assessment is needed to inform mitigation options (Robertson and Saad, 2019; Grizzetti et al., 2021). Using such maps, fertiliser management to increase nutrient use efficiency and in-field strategies (Xia et al., 2020; Martínez-Dalmau et al., 2021) could be better targeted at relatively fine scale to address problem hotspots. Furthermore, policy-makers can identify which regions are more spatially homogenous in export patterns, where broad actions could be applied, and which areas are more heterogenous with more localised actions needed. However, the implication and novelty of our study goes beyond these maps. Model predictions have inherent uncertainties that cause errors, as they are approximations of reality described by a limited set of obtainable parameters. Rather than avoid using models because of a perceived lack of credibility – a major reason for the implementation gap between research and its incorporation into policy and decision making (Willcock et al., 2016; Dubois et al., 2020) – understanding causes of uncertainties would allow communication of confidence over a modelled area and the limits to available approaches. Validation is key for such an assessment (Bryant et al., 2018; Willcock et al., 2023). We indicate several directions that would improve nutrient export and erosion modelling, which includes sub-country databases on fertiliser and manure usage, enhanced mapped differentiation of grassland types including stocking density categories, low flow nutrient measurements, as well as openly available fine scale representation of AES options for future correlations, *e.g.,* collated per country on a NUTS level 3.

CRediT authorship contribution statement

Danny A.P. Hooftman: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Data curation, Conceptualization. **Guy Ziv:** Writing – review & editing, Funding acquisition, Conceptualization. **Paul M. Evans:** Writing – review & editing, Conceptualization. **James M. Bullock:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Software and data availability

A detailed software and data availability description is given in the Supplementary Material.

Funding

We received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 817501 – BESTMAP. DAPH was contracted from UKCEH under contract 06711. The authors declare no conflict of interest.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2025.106657.

References

Addy, J.W., et al., 2022. A heteroskedastic model of park grass spring hay yields in response to weather suggests continuing yield decline with climate change in future decades. Journal of the Royal Society Interface 19, 20220361.
Bagstad, K.J., et al., 2018. The sensitivity of ecosystem service models to choices of input data and spatial resolution. Appl. Geogr. 93, 25–36.
Bagstad, K.J., et al., 2020. Towards ecosystem accounts for Rwanda: tracking 25 years of change in flows and potential supply of ecosystem services. People and Nature 2, 163–188.
Bryant, B.P., et al., 2018. Transparent and feasible uncertainty assessment adds value to applied ecosystem services modeling. Ecosyst. Serv. 33, 103–109.
Bullock, J.M., et al., 2024. Mapping the ratio of agricultural inputs to yields reveals areas with potentially less sustainable farming. Sci. Total Environ. 909, 168491.
Copernicus, 2020. Hydro river network database 2006-2012. https://doi.org/10.2909/393359a7-7ebd-4a52-80ac-1a18d5f3db9c.
de Groot, R., et al., 2012. Global estimates of the value of ecosystems and their services in monetary units. Ecosyst. Serv. 1, 50–61.
de Vries, W., 2021. Impacts of nitrogen emissions on ecosystems and human health: a mini review. Current Opinion in Environmental Science & Health 21, 100249.
de Vries, W., et al., 2022. Impacts of nutrients and heavy metals in European agriculture. In: Current and Critical Inputs in Relation to Air, Soil and Water Quality, ETC-DI Report 2022/01. European Environment Agency.
Defew, L.H., et al., 2013. Uncertainties in estimated phosphorus loads as a function of different sampling frequencies and common calculation methods. Mar. Freshw. Res. 64, 373–386.
Dubois, N.S., et al., 2020. Bridging the research-implementation gap requires engagement from practitioners. Conservation Science and Practice 2, e134.
d'Andrimont, R., et al., 2021. From parcel to continental scale–A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. Remote Sensing of Environment 266, 112708.
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. J. Hydrol. 350, 144–153.
Eigenbrod, F., et al., 2010. The impact of proxy-based methods on mapping the distribution of ecosystem services. J. Appl. Ecol. 47, 377–385.

Einarsson, R., et al., 2021. Crop production and nitrogen use in European cropland and grassland 1961–2019. Sci. Data 8, 288. https://www.nature.com/articles/s41597-021-01061-z.
EMEP, 2022. EMEP MSC-W modelled air concentrations and depositions. https://www.emep.int/mscw/mscw_moddata.html.
European Commission, 2018. Sustainable development in the European union — monitoring report on progress towards the SDGs in an EU context - 2018 edition. https://ec.europa.eu/eurostat/web/products-flagship-publications/w/ks-04-23-184.
European Environment Agency, 2020. Concentrations of heavy metals and nutrients in agricultural soils. https://www.eea.europa.eu/en/datahub/datahubitem-view/edbbd466-b845-4e4c-acf9-905ec5e28766?activeAccordion=1083478%2C1083479.
European Environment Agency, 2022. Waterbase - Water quality ICM. https://sdi.eea.europa.eu/data/bdeadea2-cfaf-4724-b002-816d71c7e361.
European Space Agency, 2020. ESA WorldCover 2020. https://worldcover2020.esa.int/downloader.
FAO, 2015. International boundaries polygons level 0 – gaul. https://datacore-gn.unepgrid.ch/geonetwork/srv/api/records/e560f98a-f6e4-41a1-bf6b-e0c99fe75426.
Gilbert, P.M., et al., 2018. Key questions and recent research advances on harmful algal blooms in relation to nutrients and eutrophication. In: Global Ecology and Oceanography of Harmful Algal Blooms. Springer, Heidelberg, Germany, pp. 229–259.
Good, A.G., Beatty, P.H., 2011. Fertilizing nature: a tragedy of excess in the commons. PLoS Biol. 9, e1001124.
Grêt-Regamey, A., et al., 2014. On the effects of scale for ecosystem services mapping. PLoS One 9, e112601.
Grizzetti, B., et al., 2021. How EU policies could reduce nutrient pollution in European inland and coastal waters. Glob. Environ. Change 69, 102281.
Guaita-García, N., et al., 2020. Environmental scenario analysis on natural and social-ecological systems: a review of methods, approaches and applications. Sustainability 12, 7542.
Henschel, S., Chan, G., 2013. Health Risks of Air Pollution in Europe—Hrapie Project. World Health Organization, Copenhagen, Denmark.
Hooftman, D.A.P., et al., 2022. Reducing uncertainty in ecosystem service modelling through weighted ensembles. Ecosyst. Serv. 53, 101398.
Huang, Y., Hsu, J.C., 2007. Hochberg's step-up method: cutting corners off Holm's step-down method. Biometrika 94, 965–975.
Huntingford, C., et al., 2022. Nitrogen cycle impacts on CO_2 fertilisation and climate forcing of land carbon stores. Environ. Res. Lett. 17, 044072.
Lamy, T., et al., 2016. Landscape structure affects the provision of multiple ecosystem services. Environ. Res. Lett. 11, 124017.
Lavorel, S., et al., 2022. Templates for multifunctional landscape design. Landsc. Ecol. 37, 913–934.
Li, Y.H., et al., 2017. Fate of nitrogen in subsurface infiltration system for treating secondary effluent. Water Sci. Eng. 10, 217–224.
Lloyd, C.T., et al., 2019. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. Big Earth Data 3, 108–139.
Ludemann, C.I., et al., 2022. Global data on fertilizer use by crop and by country. Sci. Data 9, 501. https://doi.org/10.5061/dryad.2rbnzs7qh. https://datadryad.org/stash/dataset/.
Maes, J., et al., 2012. Mapping ecosystem services for policy support and decision making in the european union. Ecosyst. Serv. 1, 31–39.
Mandle, L., Batista, N.M., 2024. Database of publications using InVEST and other natural capital project software. https://purl.stanford.edu/bb284rg5424.
Marston, C.G., et al., 2023. LCM2021–the UK land cover map 2021. Earth Syst. Sci. Data Discuss. 2023, 1–35.
Martínez-Dalmau, J., et al., 2021. Nitrogen fertilization. A review of the risks associated with the inefficiency of its use and policy responses. Sustainability 13, 5625.
Martínez-López, J., et al., 2019. Towards globally customizable ecosystem service models. Sci. Total Environ. 650, 2325–2336.
Midolo, G., et al., 2019. Impacts of nitrogen addition on plant species richness and abundance: a global meta-analysis. Global Ecol. Biogeogr. 28, 398–413.
Oenema, O., et al., 2021. Driving forces of farming systems and their impacts on CNP ratios and flows. Nutri2Cycle. https://literatur.thuenen.de/digbib_extern/dn067062.pdf.
Panagos, P., et al., 2022. Improving the phosphorus budget of European agricultural soils. Sci. Total Environ. 853, 158706.
Pascual, U., et al., 2017. Valuing nature's contributions to people: the IPBES approach. Curr. Opin. Environ. Sustain. 26, 7–16.
Pereira, P., et al., 2025. Ecosystem services mapping and modelling. Where is the validation? Geogr. Sustain. 6, 100286.
Prestele, R., et al., 2016. Hotspots of uncertainty in land-use and land-cover change projections: a global-scale model comparison. Glob. Change Biol. 22, 3967–3983.
Qiu, J., et al., 2021. Land-use intensity mediates ecosystem service tradeoffs across regional social-ecological systems. Ecosystems and People 17, 264–278.
Redhead, J.W., et al., 2018. National scale evaluation of the InVEST nutrient retention model in the United Kingdom. Sci. Total Environ. 610, 666–677.
Robertson, D.M., Saad, D.A., 2019. Spatially Referenced Models of Streamflow and Nitrogen, Phosphorus, and suspended-sediment Loads in Streams of the Midwestern United States. US Geological Survey. No. 2019-5114.
Ross, C.W., et al., 2018. Global Hydrologic Soil Groups (Hysogs250M) for Curve Number-based Runoff Modeling. ORNL DAAC, Oak Ridge, Tennessee, USA.
Schoukens, H., 2017. Nitrogen deposition, habitat restoration and the EU habitats directive: moving beyond the deadlock with the Dutch programmatic nitrogen approach? Biol. Conserv. 212, 484–492.
Schuwirth, N., et al., 2019. How to make ecological models useful for environmental management. Ecol. Model. 411, 108784.
Sharp, R., et al., 2020. Invest 3.8.7 User's Guide. The Natural Capital Project, Stanford, USA.
Tian, D., Niu, S., 2015. A global analysis of soil acidification caused by nitrogen addition. Environ. Res. Lett. 10, 024019.
UN Environment Programme, 2019. Frontiers 2018/19 Emerging Issues of Environmental Concern. United Nations Environment Programme, Nairobi, Kenya.
van der Knijff, J.M., et al., 2000. Soil erosion risk assessment in Europe. https://www.unisdr.org/files/1581_ereurnew2.pdf.
van Grinsven, H.J.M., et al., 2014. Nitrogen use and food production in European regions from a global perspective. J. Agric. Sci. 152, 9–19.
Vereecken, H., et al., 2016. Modeling soil processes: review, key challenges, and new perspectives. Vadose Zone J. 15 vzj2015.09.0131.
von Schiller, D., et al., 2008. Inter-annual, annual, and seasonal variation of P and N retention in a perennial and an intermittent stream. Ecosystems 11, 670–687.
Wang, J., et al., 2018. Global analysis of agricultural soil denitrification in response to fertilizer nitrogen. Sci. Total Environ. 616, 908–917.
Willcock, S., et al., 2016. Do ecosystem service maps and models meet stakeholders' needs? A preliminary survey across Sub-Saharan Africa. Ecosyst. Serv. 18, 110–117.
Willcock, S., et al., 2019. A continental-scale validation of ecosystem service models. Ecosystems 22, 1902–1917.
Willcock, S., et al., 2020. Ensembles of ecosystem service models can improve accuracy and indicate uncertainty. Sci. Total Environ. 747, 141006.
Willcock, S., et al., 2023. Model ensembles of ecosystem services fill global certainty and capacity gaps. Sci. Adv. 9, eadf5492.
Xia, Y., et al., 2020. Recent advances in control technologies for non-point source pollution with nitrogen and phosphorous from agricultural runoff: current practices and future prospects. Applied Biological Chemistry 63, 1–13.
Zawadzka, J., et al., 2019. Ecosystem services from combined natural and engineered water and wastewater treatment systems: going beyond water quality enhancement. Ecol. Eng. 142, 100006.