Assessing nitrate groundwater hotspots in Europe reveals a severely inadequate designation of Nitrate Vulnerable Zones

J Serra[1], C Marques-dos-Santos[1], J Marinheiro[1], S Cruz[1], MR Cameira[2], W de Vries[3], T Dalgaard[4], NJ Hutchings[4], M Graversgaard[4], F Giannini-Kurina[4], L Lassaletta[5], A Sanz-Cobena[5], M Quemada[5], E Aguilera[5], S Medinets[6,7], R Einarsson[8], J Garnier[9]

[1] Forest Research Centre CEF, Associate Laboratory TERRA, Instituto Superior de Agronomía, Universidade de Lisboa, 1349-017 Lisbon, Portugal

[2] LEAF-Linking Landscape, Environment, Agriculture and Food-Research Center, Associated Laboratory TERRA, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal;

[3] Environmental Systems Analysis Group, Wageningen University and Research, Wageningen, the Netherlands

[4] Department of Agroecology, Aarhus University, Blichers Allé 20, DK-8830 Tjele, Denmark

[5] CEICGRAM/ETSIAAB, Universidad Politécnica de Madrid, 28040 Madrid

[6] Odesa National I. I. Mechnikov University, 7 Mayakovskogo lane, 65082 Odesa, Ukraine

[7] UK Centre for Ecology & Hydrology (Edinburgh), Bush Estate, EH26 0QB Penicuik, UK

[8] Department of Energy and Technology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[9] SU CNRS EPHE, UMR Metis 7619, Paris, France

Supplementary material

Supplementary information

1. Predictors/covariates used in the random forest model

Table S1. List of explanatory variables used in the random forest (RF) model.

Category	Variable	Unit	Source
Hydrology	Flow accumulation	upstream cells cell^{-1}	HydroSheds
	Flow direction	(-)	
	Flow length downstream	m	
	Flow length upstream	m	
Hydrogeology	Groundwater depth	m month^{-1}	Fan et al. (2013)
	Aquifer delineation	(-)	
	Nitrate velocity	m yr^{-1}	Yang et al. (2022)
	Permeability	m^2	Huscroft et al. (2018)
	Porosity	%	
	Transmissivity	m^3 s^{-1}	
	Aquifer thickness	m	Sutanudjaja et al. (2018)
Soil properties	Bulk density 0-200 cm$^+$	x 10 kg m^{-3}	OpenLandMap
	Clay fraction 0-200 cm$^+$	%	
	Sand fraction 0 -200 cm$^+$	%	
	Soil pH 0-200 cm$^+$	(-)	
	Soil organic carbon 0-200cm$^+$	%	
	USDA soil texture 0-200$^+$ cm	(-)	
	Soil water content 0-200 $^+$cm	%	
Topography	Slope	Radians	
	Elevation	m	
	Hillshade	(-)	
	Topographic Wetness Index	(-)	
Agricultural management	N surplus 1992-2019 *	kg N ha-^{1} yr^{-1}	Batool et al. (2022)
	Crop- and grassland inorganic N inputs (NH_4 and NO_3^-) from manure and synthetic fertilisers 1961-2019	kg N ha-^{1} yr^{-1}	Tian et al. (2022)
Land use	Land use 1992-2019 **	%	Hoffman et al. (2022)
Climatic	Monthly runoff, potential evapotranspiration, precipitation, minimum and maximum temperature 1992-2019	mm and ºC	TerraClimate

$^+$ Horizons 0-10, 10-30, 30-60, 60-100, 100-200cm * Average of the 16 uncertainty models ** 2015-2019 annual land use maps are the average of each scenario

References

Batool, M., Sarrazin, F.J., Attinger, S. *et al.* (2022). Long-term annual soil nitrogen surplus across Europe (1850–2019). *Sci Data* **9**, 612. https://doi.org/10.1038/s41597-022-01693-9

Hoffmann, P., Reinhart,V., Rechid, D. **(**2022**)**. LUCAS LUC historical land use and land cover change dataset for Europe (Version 1.1). World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.26050/WDCC/LUC_hist_EU_v1.1

Huscroft, J., Gleeson, T., Hartmann, J., Börker, J. (2018). Compiling and mapping global permeability of the unconsolidated and consolidated Earth: GLobal HYdrogeology MaPS 2.0 (GLHYMPS 2.0). [Supporting Data], https://doi.org/10.5683/SP2/TTJNIU, Borealis, V1

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., and Bierkens, M. F. P. (2018). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev., 11, 2429–2453, https://doi.org/10.5194/gmd-11-2429-2018.

Tian, H., Bian, Z., Shi, H., Qin, X., Pan, N., Lu, C., Pan, S., Tubiello, F. N., Chang, J., Conchedda, G., Liu, J., Mueller, N., Nishina, K., Xu, R., Yang, J., You, L., Zhang, B. (2022). History of anthropogenic Nitrogen inputs (HaNi) to the terrestrial biosphere: a 5 arcmin resolution annual dataset from 1860 to 2019, Earth Syst. Sci. Data, 14, 4551–4568, https://doi.org/10.5194/essd-14-4551-2022.

Yang, C., Wang, L., Chen, S. *et al.* (2022). Nitrate transport velocity data in the global unsaturated zones. *Sci Data* **9**, 613. https://doi.org/10.1038/s41597-022-01621-x

Fan, Y., Li, H., Miguez-Macho, G. (2013). Global patterns of groundwater table depth. Science 339, 940–943. https://doi.org/10.1126/science.1229881

Supplementary figures

1. Annual nitrate concentrations

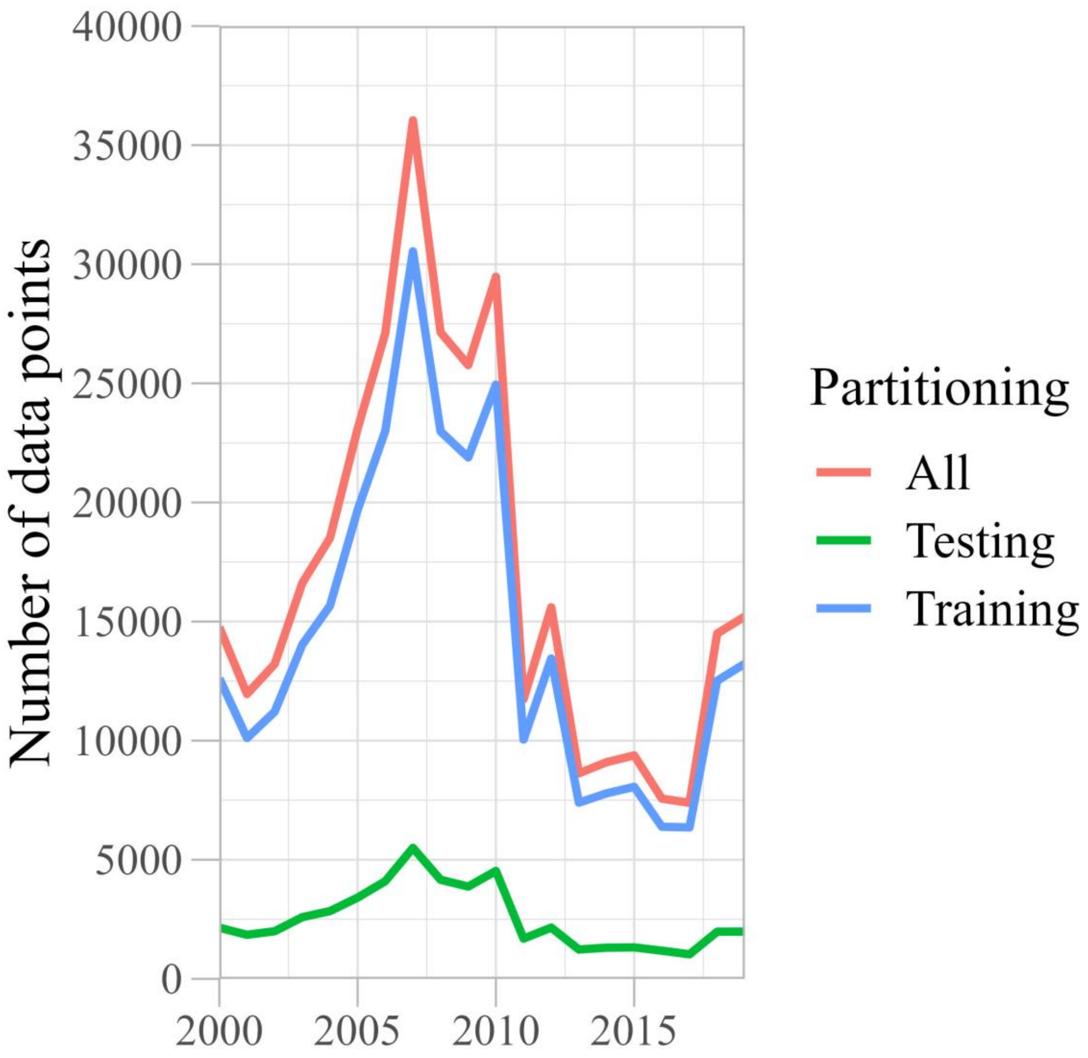

Figure S1. Number of data points used in the modelling of annual averaged nitrate concentrations for the period 2000-2019 and different partitions.

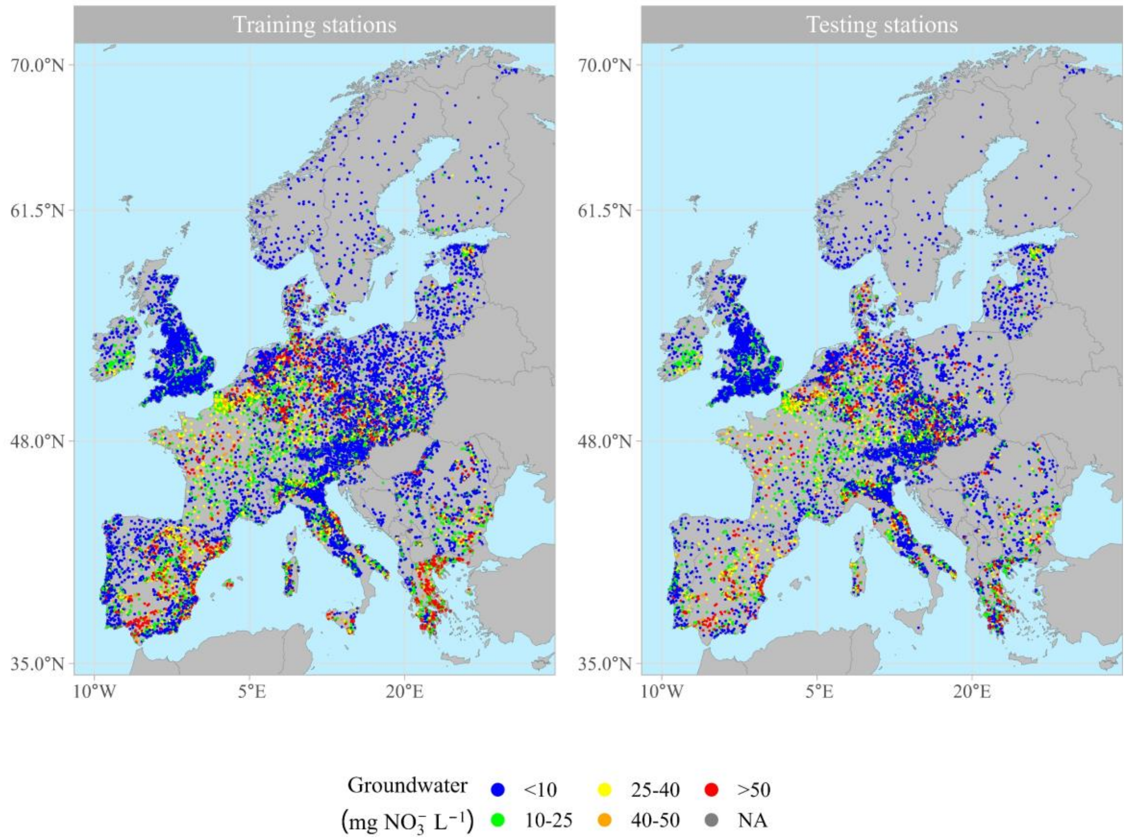

Figure S2. Spatial distribution of the nitrate monitoring stations used in the training and testing partitions for the modelling of annual nitrate concentrations.

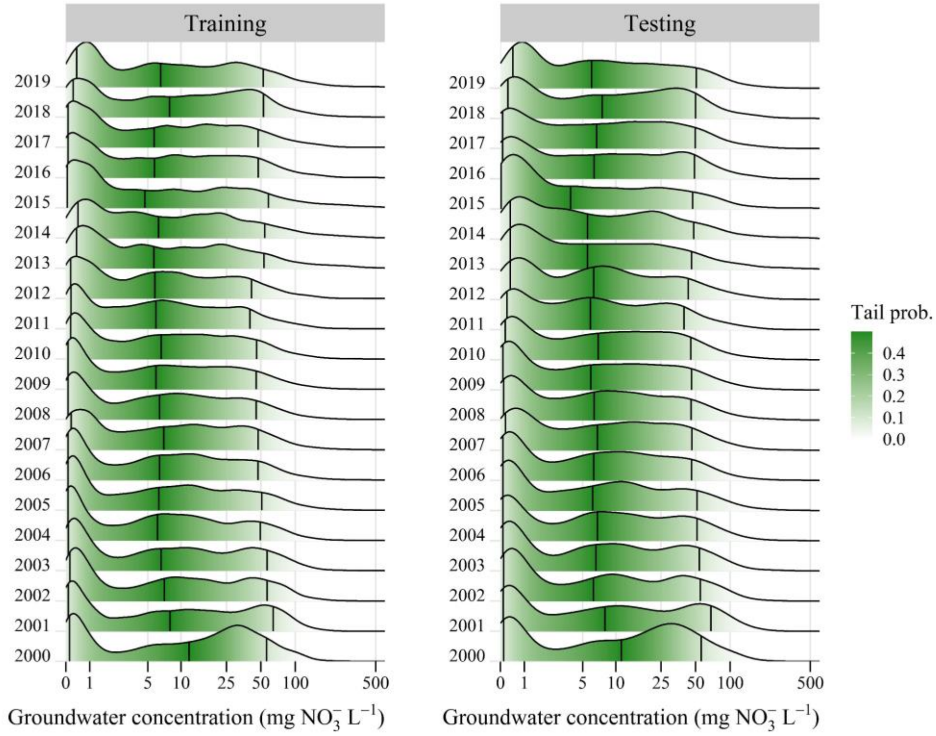

Figure S3. Annual distribution of nitrate concentrations in groundwater from the monitoring stations in the training and testing partitions.

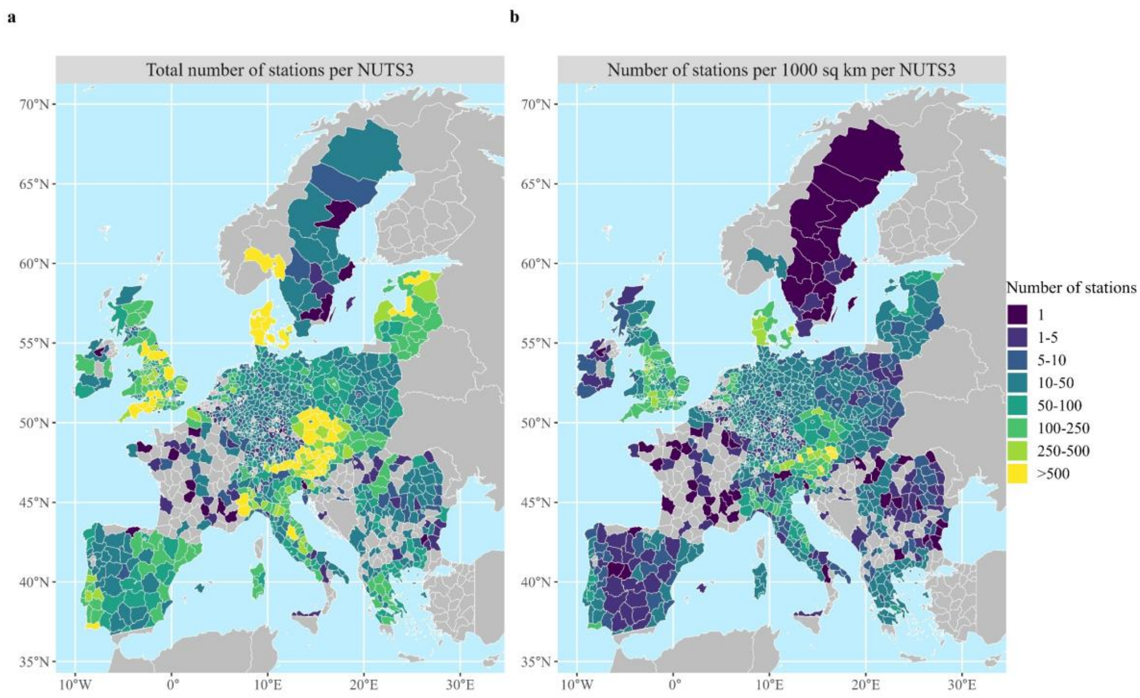

Figure S4. Number of total stations (training plus testing partitions) (a) and density relative to NUTS3 area (b), used from the EEA Waterbase (2000-2019).

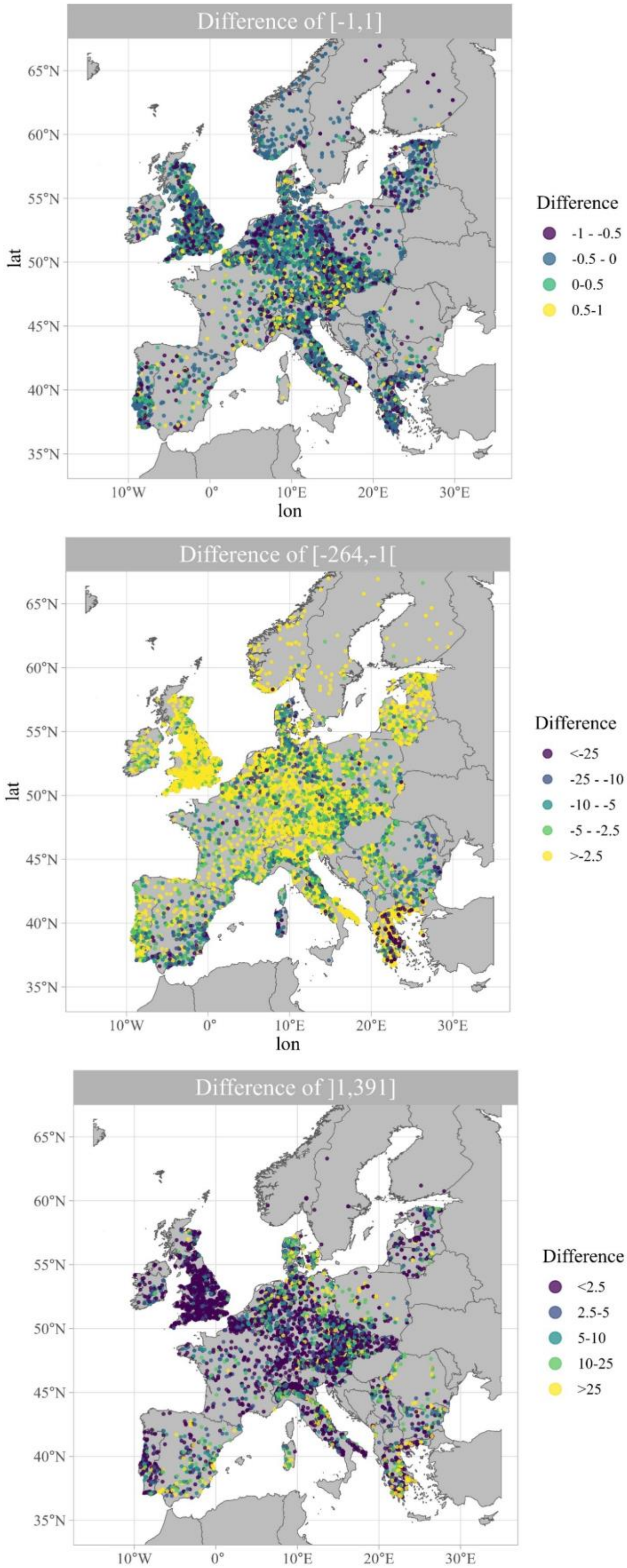

Figure S5. Difference of the observations and predictions for the testing set for the period 2000-2019. The maps represent stations with a small difference, over-estimations and under-estimations, respectively.

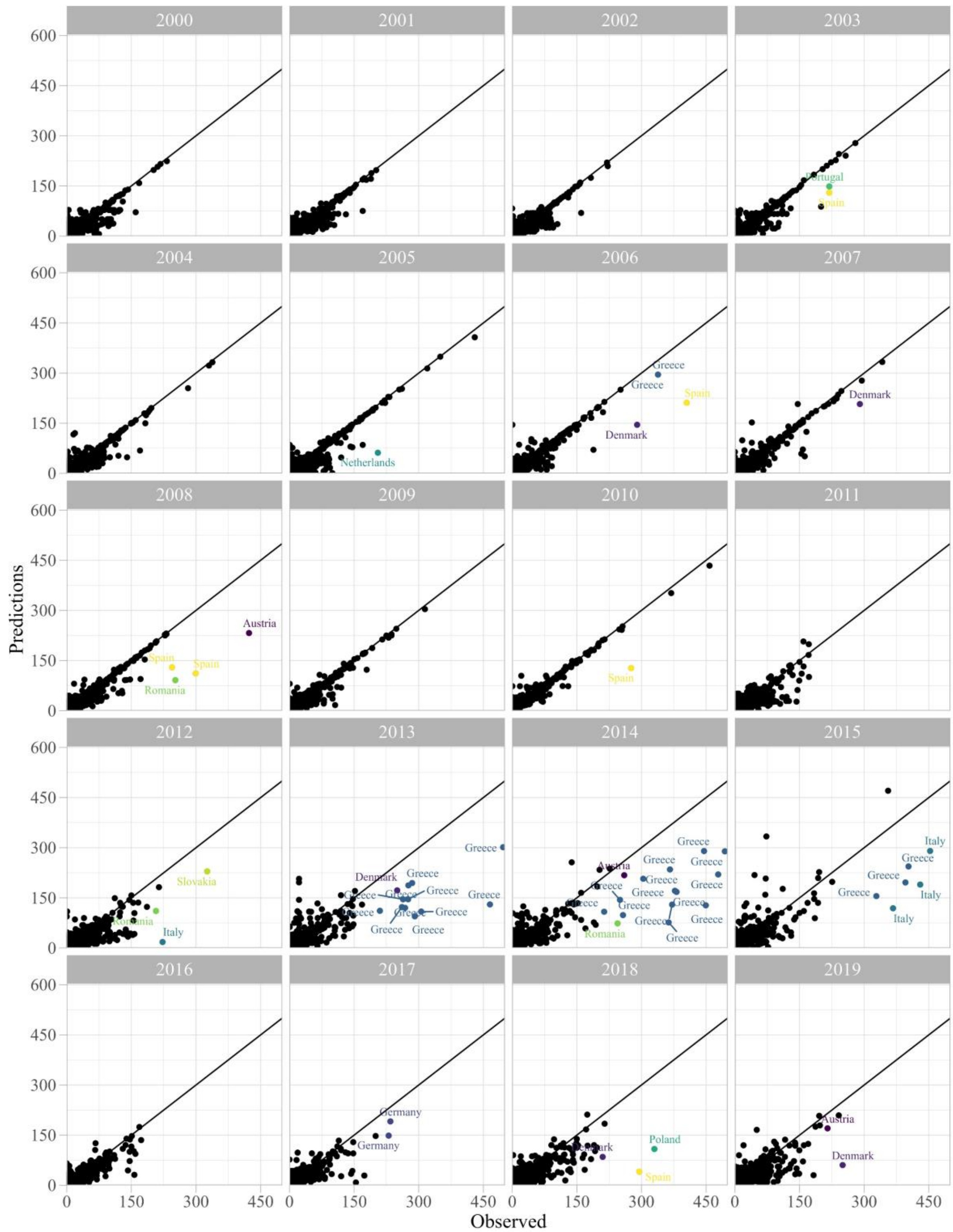

Figure S6. Scatterplot of the observed (in the testing partition) and predicted NO_3^- concentrations in groundwater for each year in the period 2000-2019. Points where the observed concentrations are (i) higher than 200 mg L^{-1} and (ii) the difference to the predictions exceed 35 mg L^{-1}. Underpredicted that for 2014 and 2015 shows it was mainly in Greece (2014) and Greece and Italy (2015).

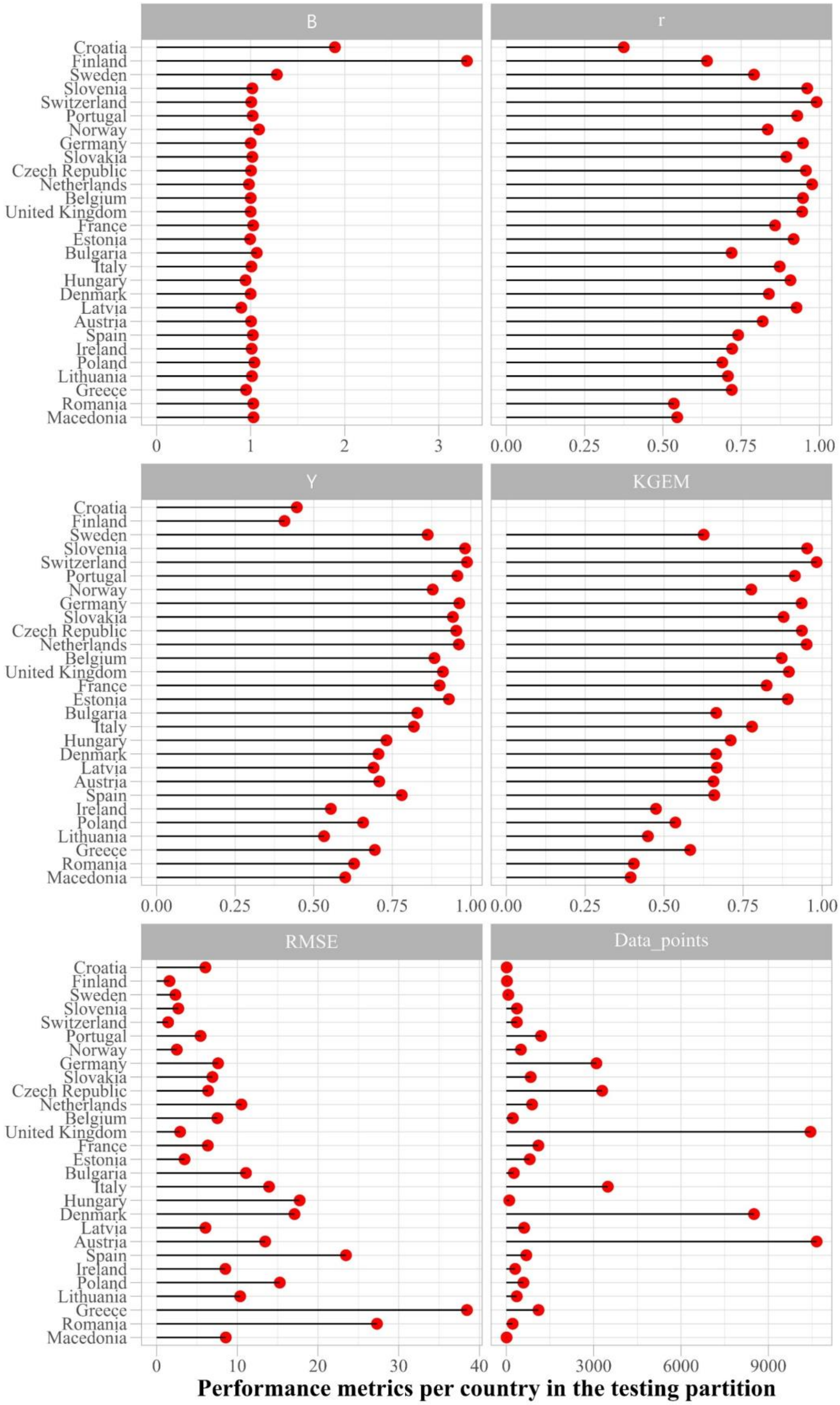

Figure S7. Performance metrics for each country in the testing partition.

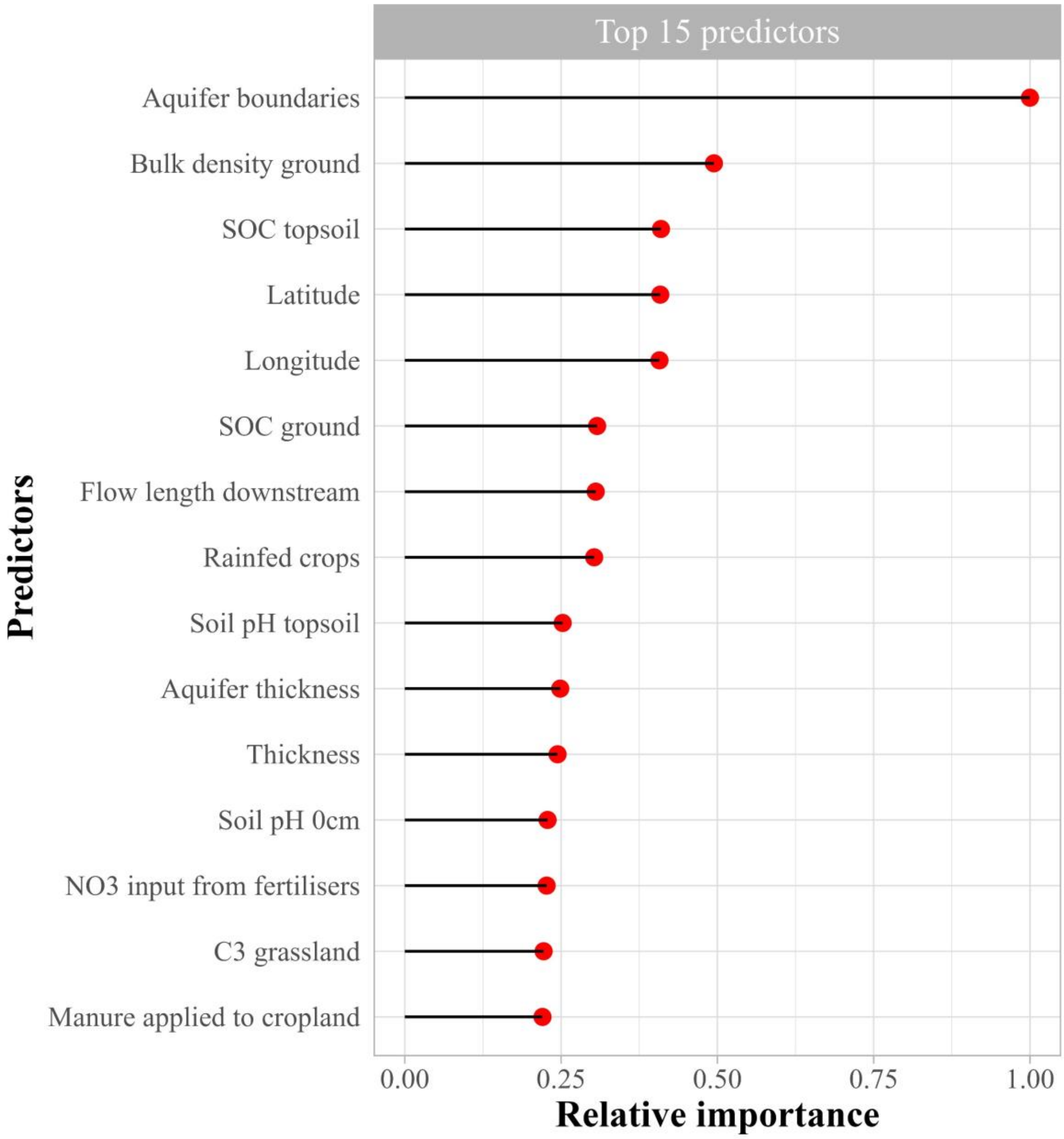

Figure S8. Top 15 most important predictors in the final random forest model for annual nitrate concentrations.

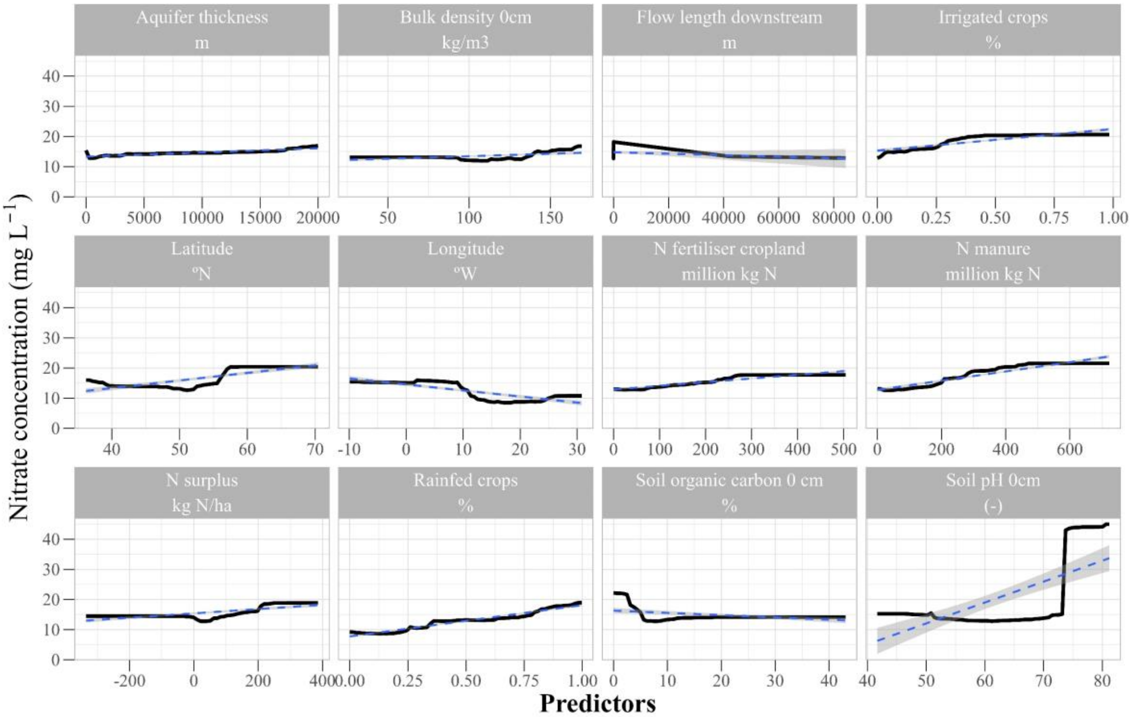

Figure S9. Partial dependence plot for selected predictors in the random forest model.

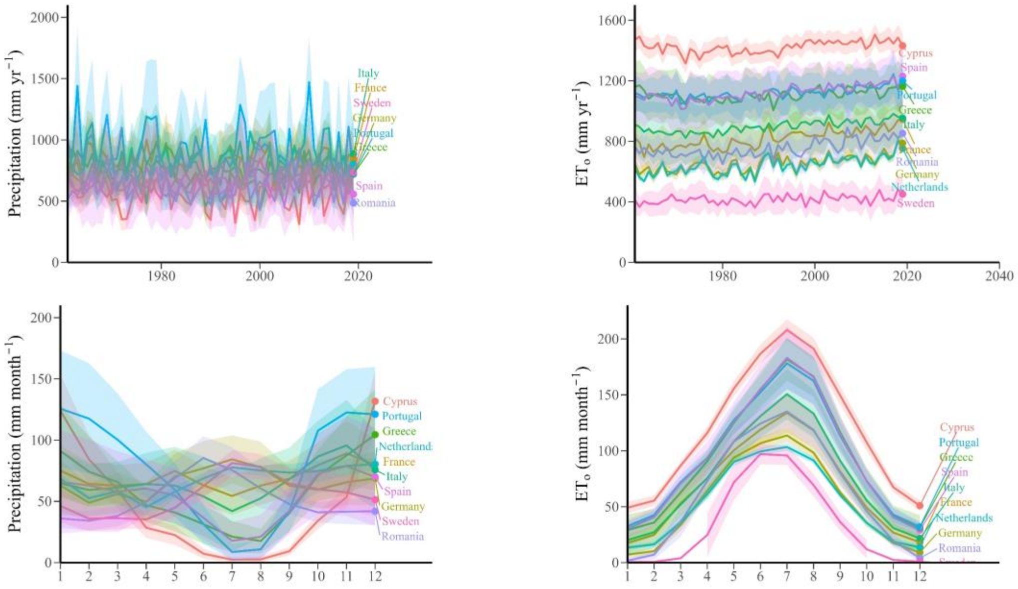

Figure S10. Annual (top panel) and monthly (bottom panel) precipitation and reference evapotranspiration for selected European countries during 1961-2019.

2. Auxiliary plots to the main text

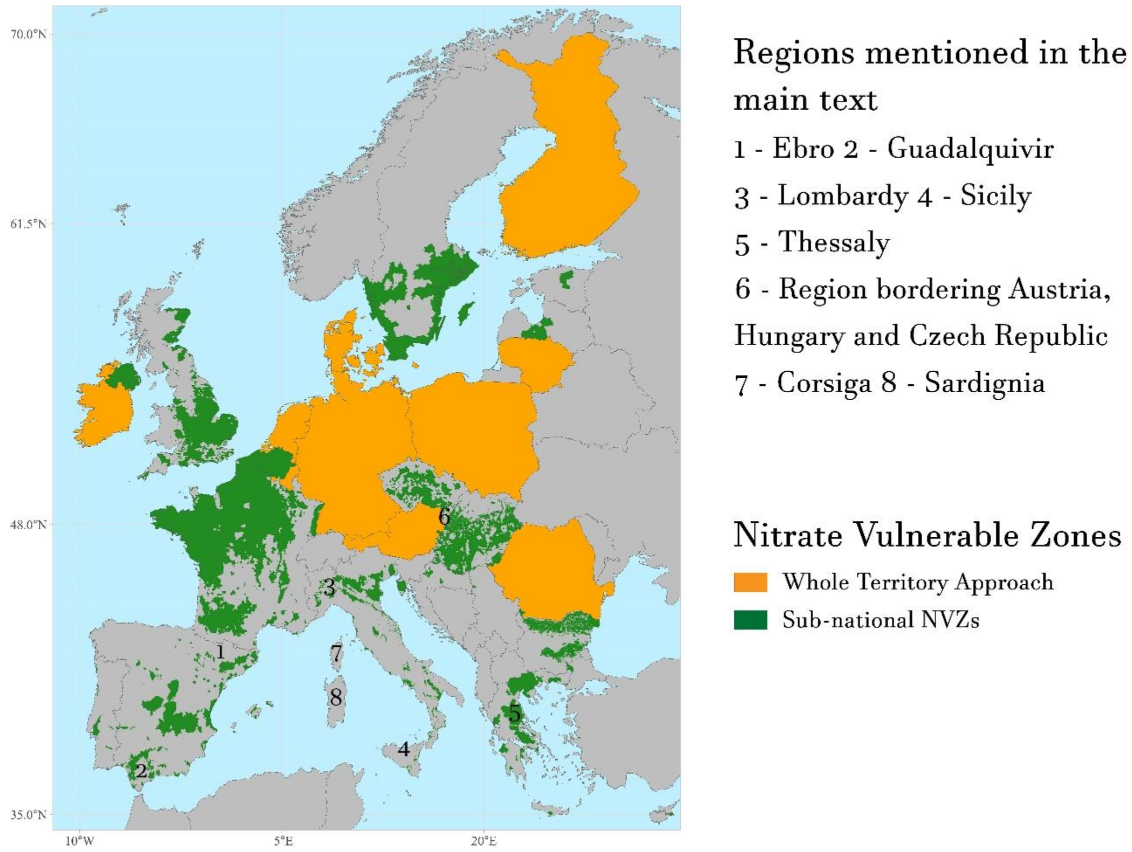

Figure S11. Spatial coverage of the Nitrate Vulnerable Zones across Europe, including some of the regions highlighted through the main text.

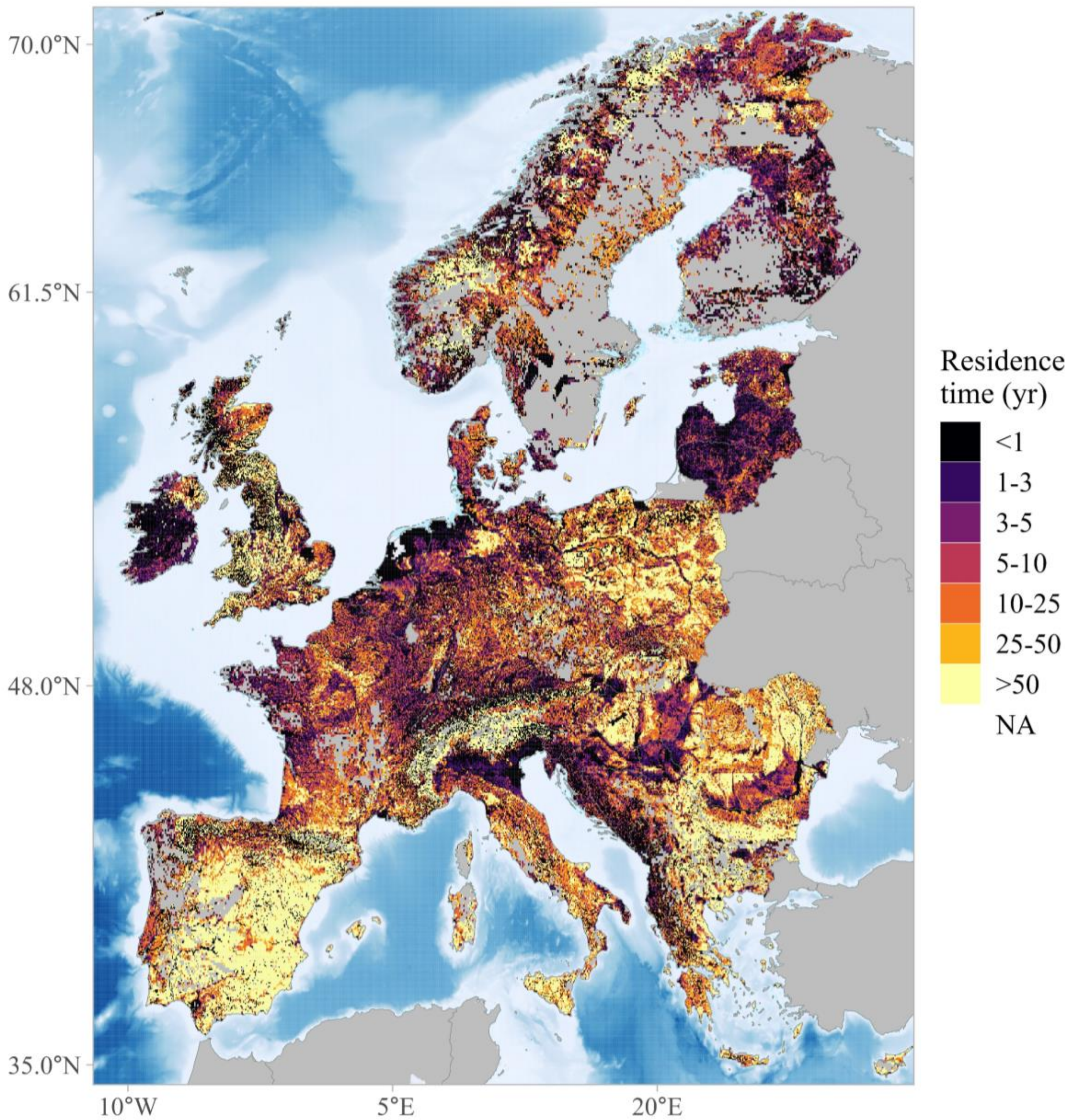

Figure S12. (Water and nitrate) Residence times in Europe, calculated using mean groundwater table depth and mean nitrate velocity from Table S1. .

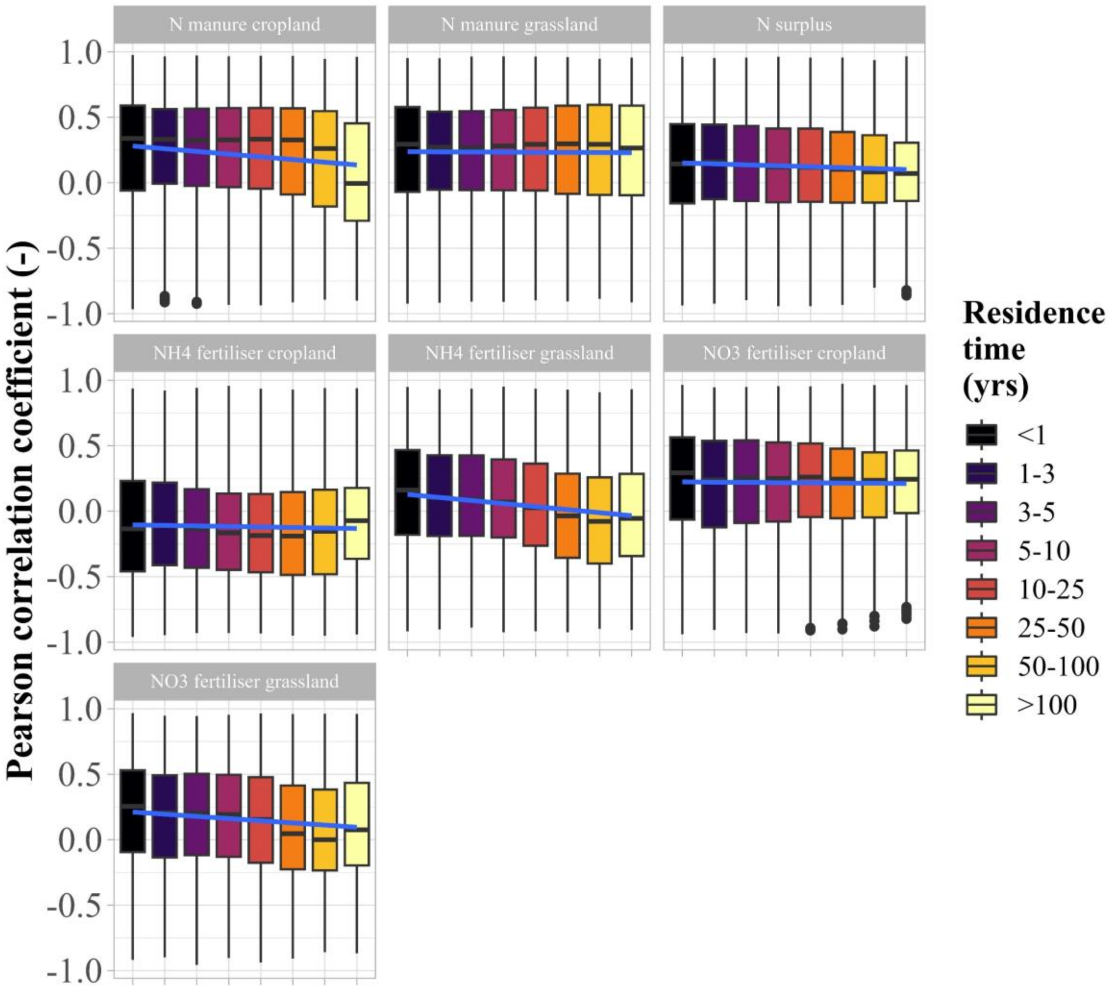

Figure S*13*. Variation of the spearman's correlation coefficient (Rho) between the different annual N predictors and nitrate concentration in groundwater, with the residence time in Europe.

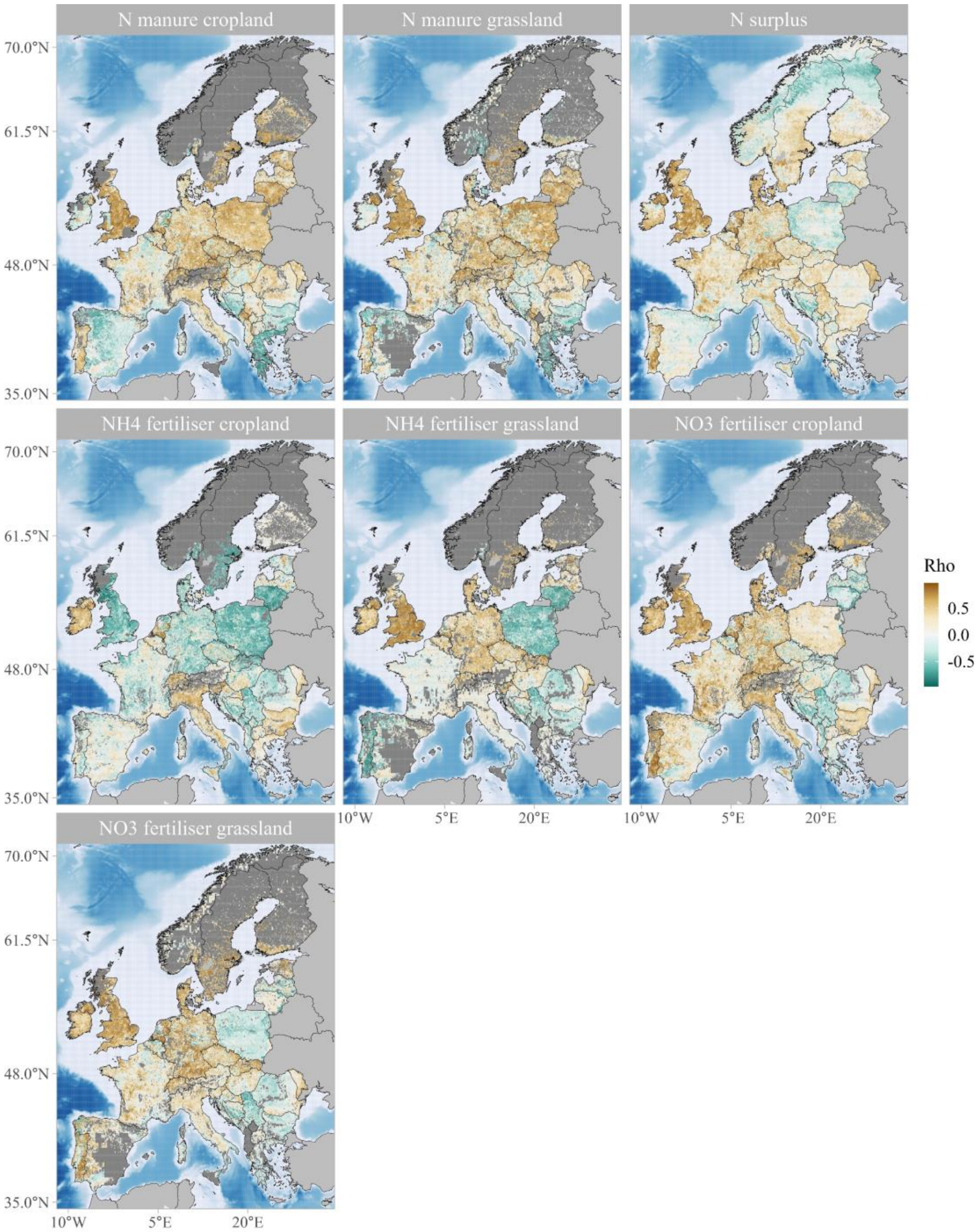

Figure S14. Spearman's correlation coefficient (Rho) between the different annual N predictors and the nitrate concentration in groundwater for the period 1992-2019. Regions in green/brown represent a high degree of responsiveness of NO_3^- concentrations with the different N predictors. That is, an increase/decrease in the N predictor resulted in a similar increase of NO_3^- concentrations in a given cell.

Table S2. Summary of the linear model (Fig. S13) employed using the Pearson correlation coefficient between nitrate concentrations and the selected N predictors, and binned residence time. Those results show non-linear relationships where statistically significant intercepts and slopes were found for the Pearson correlation coefficient and residence time. All slopes were negative, suggesting stronger monotonic relationship in regions with lower residence times.

	N surplus	**NH_4^+ fertiliser crop**	**NO_3^- fertiliser crop**	**NH_4^+ fertiliser grass**	**NO_3^- fertiliser grass**	**N manure crop**	**N manure grass**
Intercept	**0.158*** **	**-0.100*** **	**0.225*** **	**0.152*** **	**0.228*** **	**0.301*** **	**0.238*** **
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Slope	**- 0.007 *** **	**-0.004 *** **	**-0.002 *** **	**-0.023 *** **	**-0.017 *** **	**-0.021 *** **	**-0.001 ** **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Num.Obs.	295314	217469	217469	179791	200307	212553	200865
R2	0.002	0.001	0.000	0.017	0.011	0.016	0.000
R2 Adj.	0.002	0.001	0.000	0.017	0.011	0.016	0.000
AIC	238904	218189.0	194587.8	197784.3	188430.0	223004.9	201093.7
BIC	238936	218219.9	194618.6	197814.6	188460.6	223035.7	201124.3
Log.Lik.	-119449	-109091	-97290	-98889	-94211	-111499	-100543
RMSE	0.36	0.40	0.38	0.42	0.39	0.41	0.40

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

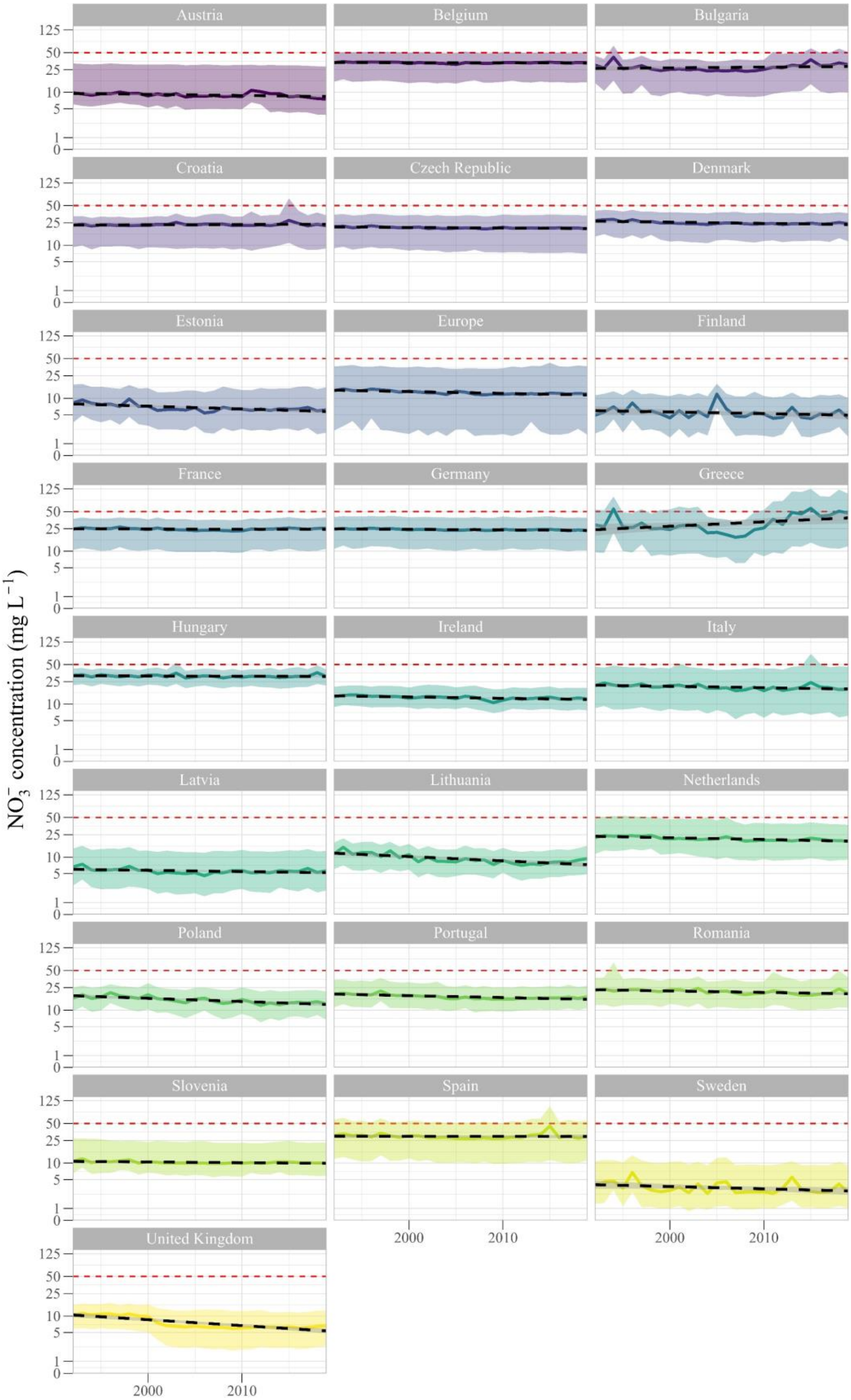

Figure S15. Country-wide trajectories of the median nitrate concentration in groundwater for the period 1992-2019. The 10th and 90th percentiles are displayed in colour as well as the slope (dashed lines).

Supplementary notes

1. Strengths and limitations

We developed a RF model that showed a good performance for Europe at a fine resolution, covering a long historical period (1992-2019). Those data are at a sufficiently fine resolution to enable the identification of hotspots inside and outside the designated NVZs as well as their trends. This can be used to forecast NO_3^- concentrations provided the model is fed with future data about N inputs, N surplus, land use and monthly climate inputs. Although we did not account for the different timelines for the designation and/or expansion of the NVZs, we provide what can be the backbone for more detailed analysis for individual NVZs. Alternatively, our model can be used to review the spatial boundaries of the NVZs to better include currently unaccounted hotspots and to reveal trends in regions without a strong monitoring network.

However, our approach is not without limitations and caveats. The model tackles the time-lag of NO_3^- contamination directly through four predictors (groundwater depth, NO_3^- velocity, porosity and year) and indirectly by allowing a detailed trend analysis in space and time. However, we are unable to directly measure the averaged time-lag of NO_3^- at the grid level (e.g., Ascott et al., 2017) due to the black-box nature of the RF model. The predictions depend, to a large extent, on the accuracy of the predictors and the EEA dataset monitoring stations.We are unable to vouch for the accuracy of N related spatial indicators for older years (e.g., 1961-1990) since they rely on extrapolations and assumptions we are unable to verify due to lack of data. Our predictions for older years at the grid level should thus be used with caution although only the NO_3^- from fertilisers applied to cropland was in the top 15 of the most important predictors. Furthermore, there are large gaps in the spatial coverage of the monitoring stations used to train the model and a disparity in the density of the national monitoring networks, which was generally higher in NW (England and Denmark) and Central (Austria, Czech Republic) Europe. The current Waterbase dataset is a commendable and valuable effort but there are some obstacles that need to be overcome for similar studies to further explore NO_3^- hotspots. We were also unable to account for multi-layered aquifers (e.g., the aquifers in the Tagus NVZ; Cameira et al., 2021) since sampling depths were not included.

References

Ascott, M. J., Gooddy, D. C., Wang, L., Stuart, M. E., Lewis, M. A., Ward, R. S., & Binley, A. M. (2017). Global patterns of nitrate storage in the vadose zone. *Nature Communications*, *8*(1), 1–6. https://doi.org/10.1038/s41467-017-01321-w

Cameira, M. do R., Rolim, J., Valente, F., Mesquita, M., Dragosits, U., & Cordovil, C. M. d. S. (2021). Translating the agricultural N surplus hazard into groundwater pollution risk : Implications for effectiveness of mitigation measures in nitrate vulnerable zones. *Agriculture, Ecosystems and Environment 306*(October 2020). https://doi.org/10.1016/j.agee.2020.107204

2. Comments on the most important predictors of the random forest model

The inherent black box nature of the random forest (RF) model limits our (statistical) understanding about the influence of the different predictors. But nonetheless we can speculate about the role of the most important predictors in NO_3^- concentrations in groundwater. The most important predictors are mainly soil properties (bulk density, soil pH, water content and clay fraction at different depths), topographic (elevation), land use (grasslands, rainfed and irrigated crops), agricultural management (NO_3^- fertilisers to cropland), hydrologic (flow length downstream) as well as climatic (precipitation in December).

Soil bulk density reflects soil's structural support and influences water (and NO_3^-) movement in the soil as well as soil aeration. High bulk densities restrict plants' growth and the downward transport of water and NO_3^-. It is thus a proxy of the agricultural hazard and vulnerability since it influences the extent to which NO_3^- and water can percolate. Similarly, water content (or field capacity) affect how much water the soil can hold at a given depth, influencing groundwater recharge and N losses through runoff and leaching. High values of soil pH (7-8) and clay fraction give a good proxy for favouring denitrification in the soil, which reduces NO_3^- available for leaching. Elevation is tied with groundwater table depth: in principle higher elevations are linked with higher water table depths, thus having longer residence time (Fan et al., 2013). Land use is associated with N management here: grasslands are better at retaining NO_3^- (Velthof et al., 2009) while rainfed and irrigated cropping systems are associated with high N inputs in NW and Southern Europe, respectively (e.g., Leip et al., 2011). Fertilisers containing NO_3^- are prone to N leaching losses; for instance, ammonium nitrate attained the largest emission factor for leaching (i.e., fraction of N input lost through leaching) from a recent meta-analysis (Wang et al., 2019). Water losses from percolation during winter are mainly driven by generally high precipitation with low reference evapotranspiration.

Our most important predictors are comparable to those from earlier studies in Europe (soil pH, field capacity) (Serra et al., 2023a), Portugal (field capacity, bulk density, soil texture, elevation) (Serra et al., 2023b) and Germany (land use, hydrogeologic units) (Knoll et al., 2019).

References

Serra, J., Marques-dos-Santos, C., Marinheiro, J., Aguilera, E., Lassaletta, L., Sanz-Cobeña, A., Garnier, J., Billen, G., de Vries, W., Dalgaard, T., Hutchings, N., & do Rosário Cameira, M. (2023a). Nitrogen inputs by irrigation is a missing link in the agricultural nitrogen cycle and related policies in Europe. *Science of The Total Environment*, *889*, 164249. https://doi.org/10.1016/j.scitotenv.2023.164249

Serra, J., Paredes, P., Cordovil, Cm. S., Cruz, S., Hutchings, N. J., & Cameira, M. R. (2023b). Is irrigation water an overlooked source of nitrogen in agriculture? *Agricultural Water Management*, *278*. https://doi.org/10.1016/j.agwat.2023.108147

Knoll, L., Breuer, L., & Bach, M. (2019). Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Science of the Total Environment*, *668*, 1317–1327. https://doi.org/10.1016/j.scitotenv.2019.03.045

Velthof, G.L., Oudendag, D., Witzke, H.P., Asman, W.A.H., Klimont, Z., Oenema, O., 2009. Integrated assessment of nitrogen losses from agriculture in EU-27 using MITERRA-EUROPE. *Journal of Environmental Quality* 38, 402. https://doi.org/10.2134/jeq2008.0108

Wang, Y., Ying, H., Yin, Y., Zheng, H., & Cui, Z. (2019). Estimating soil nitrate leaching of nitrogen fertilizer from global meta-analysis. *Science of the Total Environment*, *657*, 96–102. https://doi.org/10.1016/j.scitotenv.2018.12.029

Fan, Y., Li, H., Miguez-Macho, G. (2013). Global patterns of groundwater table depth. Science 339, 940–943. https://doi.org/10.1126/science.1229881

3. Towards better large spatial scale predictors

The quality of the RF predictions not only depends on the quality of the monitoring stations but also on how representative of NO_3^- dynamics the predictors are. There are still several data gaps concerning the predictors, particularly at the large spatial scale used in this study (European scale) and fine resolution (0.04ºx0.04º). Specifically, there are currently no fine spatial (≤10x10km) nor temporal (long term periods; monthly timestep) resolution for N leaching. We argue that having such data would greatly improve the estimates but otherwise we had to use proxies like the N surplus and N inputs from manure/synthetic fertilisers to crop- and grassland. In addition, developing seasonal/monthly N inputs datasets, albeit currently not done at large spatial scales, would likely increase the accuracy of seasonal predictions. Groundwater recharge is an important predictor still missing but that is of key importance in regulating the stored NO_3^- in the vadose zone and likewise in groundwater (Ascott et al., 2017). Currently, the PCR-GLOBWB model (Sutanudjaja et al., 2018) freely provides such data but only for the period 1958-2015. In addition, irrigation rates are currently missing from our predictors due to the relatively high spatial resolution (compared with available data; e.g., Khan et al., 2023). Irrigation, coupled with groundwater recharge, data may allow us to improve prediction accuracy in Mediterranean agricultural systems.

Despite the very good performance achieved by our model, we think the framework here proposed could be improved by implementing the NO_3^- stored in the vadose and its outflow to aquifers on an annual basis as additional predictors. Developing detailed N leaching and residence time maps would enable this calculation (Ascott et al., 2017).

References

Ascott, M. J., Gooddy, D. C., Wang, L., Stuart, M. E., Lewis, M. A., Ward, R. S., & Binley, A. M. (2017). Global patterns of nitrate storage in the vadose zone. *Nature Communications*, *8*(1), 1–6. https://doi.org/10.1038/s41467-017-01321-w

Khan, Z., Thompson, I., Vernon, C. R., Graham, N. T., Wild, T. B., Chen, M., 2023. Global monthly sectoral water use for 2000-2100 at 0.5º resolution across alternative futures. Scientific Data 10, 201. https://doi.org/10.1038/s41597-023-02086-2

Sutanudjaja, E. H., van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H. C., Drost, N., van der Ent, R. J., de Graaf, I. E. M., Hoch, J. M., de Jong, K., Karssenberg, D., López López, P., Peßenteiner, S., Schmitz, O., Straatsma, M. W., Vannametee, E., Wisser, D., & Bierkens, M. F. P. (2018). PCR-GLOBWB 2: A 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

4. Comment on the Nitrates Directive Report

The latest report from the Nitrates Directive for the period 2016-2019 (JRC, 2022) provides what we assume to be an extended dataset relative to the EEA Waterbase, also including confidential stations. The report provides maps for the monitoring stations as points and averaged at the NUTS2 and NUTS3 scales. Here we would like to point why comparisons with our gridded predictions and the regional maps provided in the report are unable to provide any meaningful baseline for comparison (Figs. S16-17).

Firstly, the NUTS regions are designated based on socioeconomic data and often do not conform to agricultural regions (Kros et al., 2018; Serra et al., 2019) nor, much even less, to the delimitation of aquifers. Therefore, although providing averaged maps at the NUTS scale is only useful to gain insight on possible hotspots across Europe. Even then, this is uncertain at best since NUTS regions markedly differ in size, ranging from 16 km^2 in the UK (Luton) to 105 thousand km^2 in Sweden (Norrbottens) – NUTS3. Secondly, with differences in size of the regions, a higher degree of heterogeneity in terms of NO_3^- concentrations is also expected. Averaging concentrations of the monitoring stations at this scale would dilute the heterogeneity in a NUTS3 region, effectively losing the capacity to highlight hotspots. Thirdly, our predicted concentrations should, in principle, have less heterogeneity relative to monitoring stations in the same cell (3x3km). Our predictions provide concentrations already averaged at each cell. Fourthly, the averaged concentrations from the report are based only on observations while our model was able to successfully detect otherwise uncovered hotspots.

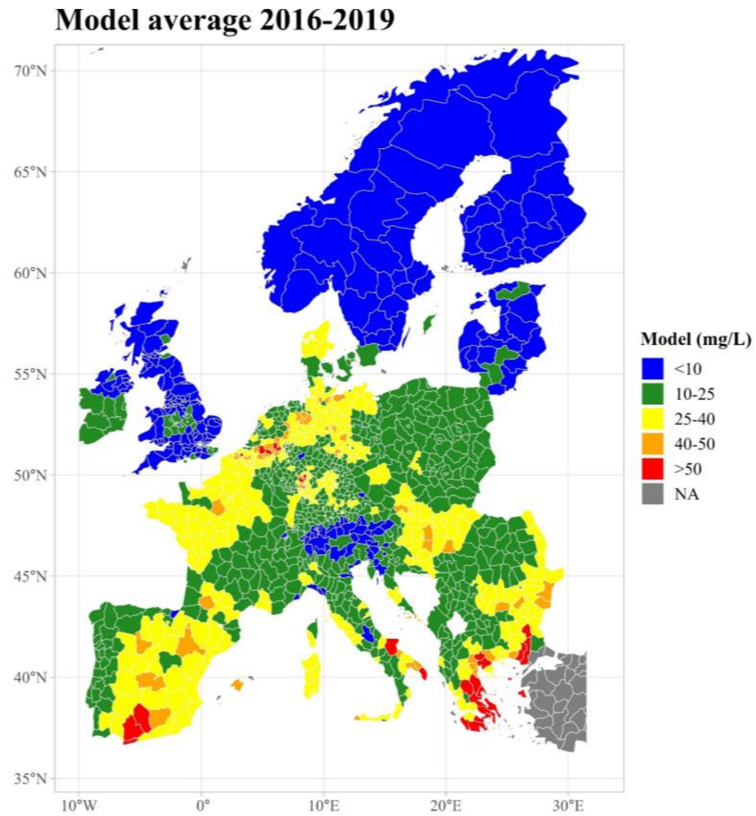

Figure S16. Averaged predictions for 2016-2019 at the NUTS3 regions across Europe.

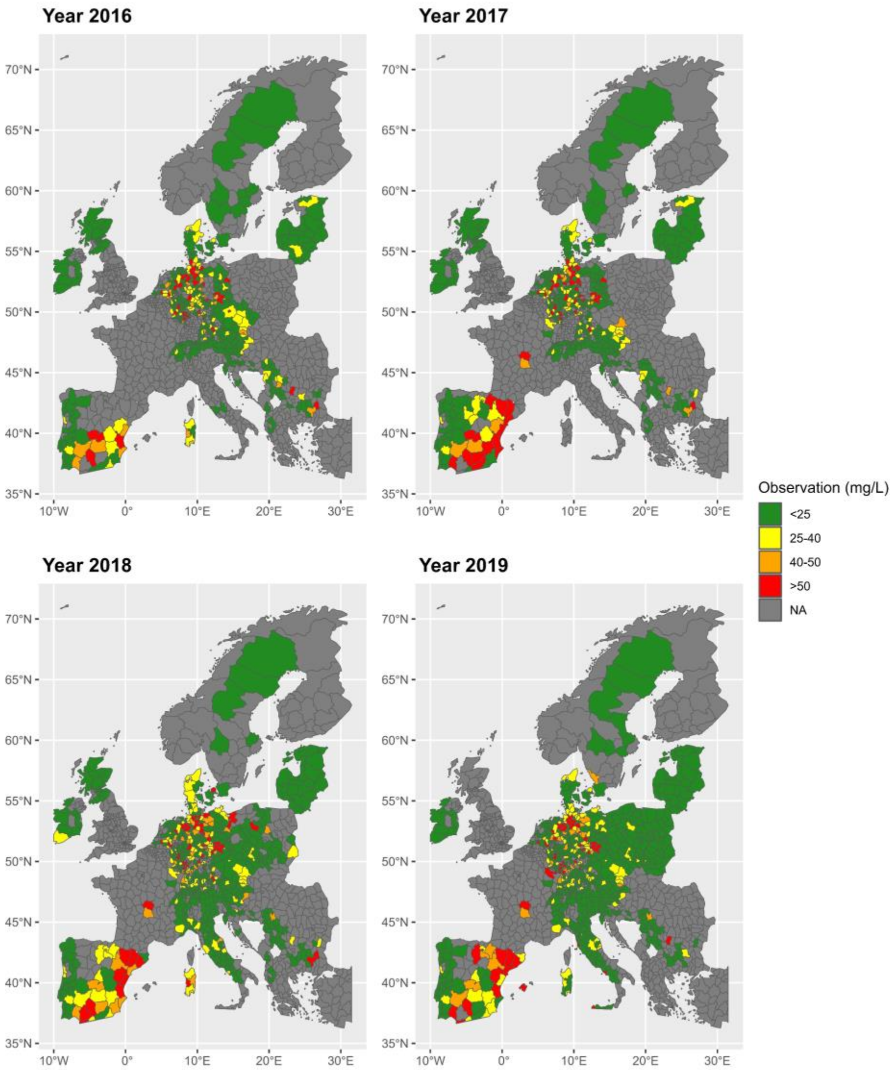

Figure S17. Annual observations from the EEA Waterbase dataset for the period 2016-2019, averaged per NUTS3 region.

References

JRC (Joint Research Center), 2022. Nitrates Directive – Reporting Period 7 (2016-2019). Dashboard accessed in 27 April 2023, https://water.jrc.ec.europa.eu/portal/apps/opsdashboard/index.html#/cb6034c2a75e4df282f8a62f90c16caa

Kros, J., Hutchings, N. J., Kristensen, I. T., Kristensen, I. S., Børgesen, C. D., Voogd, J. C., Dalgaard, T., & de Vries, W. (2018). A comparison of disaggregated nitrogen budgets for Danish agriculture using Europe-wide and national approaches. *Science of the Total Environment*, *643*, 890–901. https://doi.org/10.1016/j.scitotenv.2018.06.267

Serra, J., Cordovil, C. M. d. S., Cruz, S., & Hutchings, N. J. (2019). Challenges and solutions in identifying agricultural pollution hotspots using gross nitrogen balances. *Agriculture, Ecosystems and Environment*, *283*(June), 106568. https://doi.org/10.1016/j.agee.2019.106568

5. Trend analysis of hostpots at the national scale

Table S3. Trend analysis at the national level for Europe for three distinct periods per the spearman's correlation coefficient (p-value). Statistically significant trends ($p < 0.05$) are highlighted in bold. Spearman's correlation coefficients are coloured in red (increasing trend; rho > 0.2), light green (declining trend; rho < -0.2) and grey (stable trend).

Country	1992-2003	2004-2019	1992-2019
Portugal	-0.42 (0.17)	-0.27 (0.303)	**-0.82 (0)**
Spain	**-0.65 (0.026)**	**0.64 (0.009)**	-0.29 (0.13)
France	**0.05 (0.886)**	**0.77 (0.001)**	0.16 (0.413)
Italy	-0.07 (0.834)	**0.46 (0.074)**	-0.18 (0.367)
Czech Republic	-0.1 (0.76)	0.24 (0.372)	**0.55 (0.002)**
Sweden	NA (NA)	-0.32 (0.231)	0.12 (0.53)
Greece	0.24 (0.457)	**0.74 (0.001)**	**0.47 (0.013)**
Bulgaria	-0.18 (0.585)	**0.83 (0)**	**0.52 (0.004)**
Hungary	-0.07 (0.837)	0.26 (0.33)	0.27 (0.167)
Croatia	**0.81 (0.002)**	**0.63 (0.011)**	**0.74 (0)**
Estonia	0.52 (0.084)	0.36 (0.17)	**0.75 (0)**
Latvia	**0.67 (0.018)**	0.2 (0.461)	**0.52 (0.004)**
Slovenia	**0.59 (0.046)**	**-0.51 (0.046)**	**-0.42 (0.027)**
United Kingdom	0.17 (0.59)	**-0.83 (0)**	**-0.42 (0.026)**
Slovakia	0.26 (0.422)	**-0.67 (0.005)**	-0.12 (0.552)
Netherlands	**-0.69 (0.013)**	**-0.77 (0.001)**	**-0.89 (0)**
Germany	**0.76 (0.007)**	0.26 (0.321)	**0.8 (0)**
Finland	NA (NA)	-0.36 (0.166)	-0.01 (0.952)
Romania	-0.23 (0.477)	0.67 (0.005)	0.09 (0.637)
Belgium	0.14 (0.656)	-0.43 (0.098)	-0.72 (0)
Denmark	**-0.71 (0.01)**	**-0.57 (0.022)**	**-0.75 (0)**
Austria	0.11 (0.733)	-0.21 (0.44)	0.23 (0.231)
Ireland	0.52 (0.084)	-0.49 (0.053)	-0.1 (0.609)
Lithuania	**0.75 (0.005)**	-0.26 (0.325)	**0.42 (0.027)**
Poland	**0.6 (0.038)**	**0.5 (0.047)**	**0.77 (0)**

6. Uncertainty analysis

We performed an uncertainty analysis using the hyper tuned random forest models. We modified four predictors – those more directly linked with the Nitrates Directive: the N surplus, N input from manure, N fertiliser applied to crop- and grassland. Each predictor was modified by multiplying the annual values (modifiers) by ± 0.75, ±0.9, ± 0.95, ± 1.05, ± 1.1 and ± 1.25. We simulated the impact of each predictor individually and at the same time. We calculated the number of hotspots in each simulation at the European and national levels and compare those data with the baseline.

The hotspots size in Europe changed between -6.8 and 8.7% by varying each predictor 25% at the same time (Fig. S18). This implies that reducing these N inputs (and N surplus) by 25% would only reduce total hotspot size in Europe in 6.8%. NO_3^- fertiliser inputs to cropland had the largest impact (-3.4 – 3.7%) while the NO_3^- input to grassland had a negligible impact (-0.3 – 0.3%). The impact of N manure (-2.0 – 2%) was comparable to the N surplus (-1.2 – 2.5%).

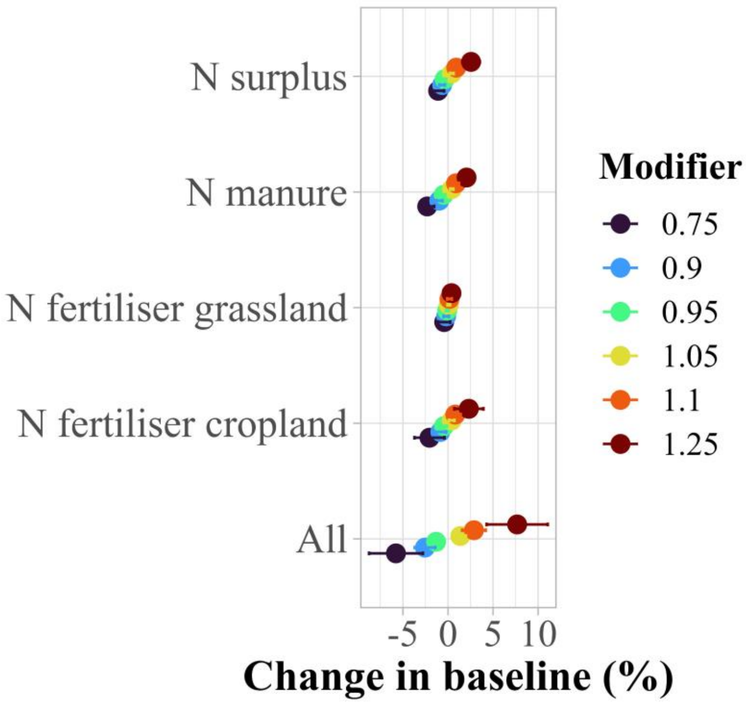

Figure S18. Uncertainty analysis of hotspot size in Europe relative to the baseline from four predictors: N surplus, N manure applied to cropland, N fertiliser input to crop- and grassland as well as their joint impact. The point represents the annual average change in baseline and the error bars represent the standard deviation.

This impact was, however, not equal across Europe (Fig. S19) neither in terms of overall impact nor in terms of predictor impact. Concerning the former, The Netherlands, Germany, Denmark, Belgium and France were the countries with the highest uncertainty to the joint modification of all N predictors. Indeed, by increasing all N predictors 20%, hotspot size in the Netherlands increased more than 50% in some years. Reducing all N predictors in 20% yielded an average reduction in hotspot size of 15%.

In terms of predictor impact, the N surplus was, by far, the most uncertain predictor in the Netherlands on which the hotspot size grew upwards to 40% by increasing the N surplus by 25%. Conversely, N inputs from manure were most sensitive in Belgium and Denmark where hotspot size increased to almost 13 and 20%, respectively, by increasing manure in 20%. Livestock production is arguably a key driver for N pollution in these countries (van Grinsven et al., 2012). The low uncertainty of N manure in the Netherlands is surprising given the high livestock density (Gilbert et al., 2018). It is unclear why this was the case, but we may speculate it is due to a combination of more efficient application, higher enforcement of treatments during disposal and/or transport to elsewhere (e.g., NW Germany). The N input from synthetic fertilisers applied in cropland was generally the most uncertain predictor across Europe (e.g., Bulgaria, France, Greece).

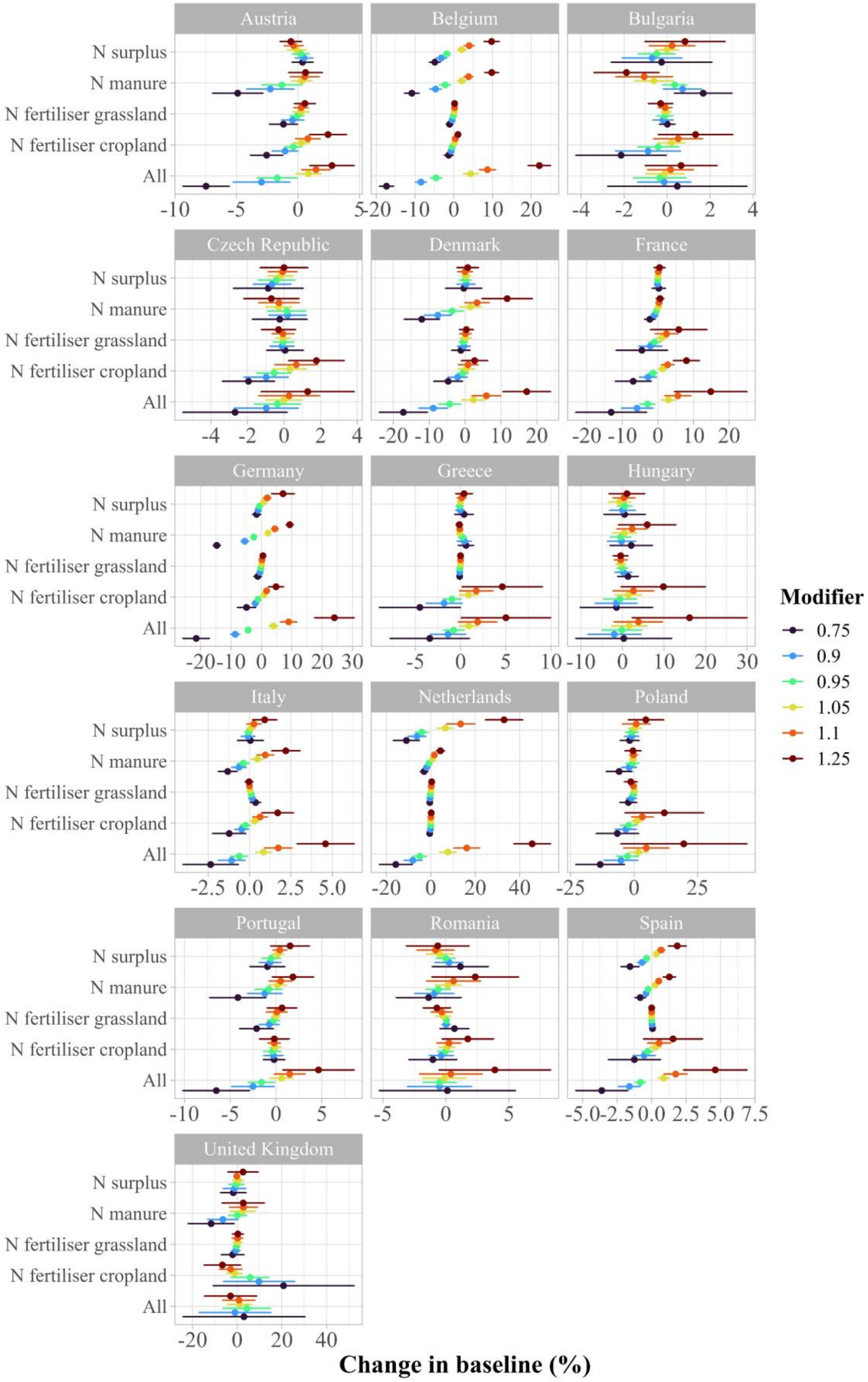

Figure S19. Uncertainty analysis of hotspot size in selected countries relative to the baseline from four predictors: N surplus, N manure applied to cropland, N fertiliser input to crop- and grassland as well as their joint impact. The point represents the annual average change in baseline and the error bars represent the standard deviation.

References

Van Grinsven, H.J.M., Ten Berge, H.F.M., Dalgaard, T., Fraters, B., Durand, P., Hart, A.,Hofman, G., Jacobsen, B.H., Lalor, S.T.J., Lesschen, J.P., Osterburg, B., Richards, K.G.,Techen, A.K., Vertès, F., Webb, J., Willems, W.J., 2012. Management, regulation andenvironmental impacts of nitrogen fertilization in northwestern Europe under the Nitrates Directive; a benchmark study. Biogeosciences 9, 5143–5160. https://doi.org/10.5194/bg-9-5143-2012.

Gilbert, M., Nicolas, G., Cinardi, G. *et al.* (2018). Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010. *Sci Data* **5**, 180227. https://doi.org/10.1038/sdata.2018.227

7. Searching for causality between hotspot trends and the Nitrates Directive

To gain insight on the possible effect of the ND in NO_3^- concentration in groundwater, we used the same hypertuned model but with varying levels of predictors. Specifically, we focussed on those more likely to be changed with the implementation of the ND: (i) N management (N inputs from synthetic fertilisers and manure to crop- and grassland, and N surplus) and land use (crop- and grassland area). Given the different but generally decreasing trajectories of N surplus since the 1990s, we decided to compare our predictions (hereby called baseline) with two other alternative scenarios: historical (1961-1990) and present years (2000-2019) conditions for these predictors. Both the historical and present scenarios here represent the average of each predictor for the years 1961-1990 and 2000-2019, respectively. We then again predicted the annual maps of NO_3^- concentrations for 1992-2019 and respective hotspot analysis for the other two scenarios (1961-1990 and 2000-2019). The hotspot analysis focussed, again, at the European and national levels.

To explore whether statistically significant differences were simulated for the two scenarios, we implemented a one-way anova test followed by a hockey test. Significant differences were reported if $p < 0.05$.

This methodological approach is insufficient to conclude whether there is causality or not in the hotspot trends related with the ND. Nonetheless, it allows us to gain insight, to some extent, of the implications of the ND implementation in hotspot size across Europe. This approach is complemented by implementing an uncertainty analysis for these predictors as well (See “6. Uncertainty analysis”).

Europe

In Europe, the largest hotspot area was attained using the historical scenario (115 ± 50 $km^2\ yr^{-1}$), baseline (107 ± 42 $km^2\ yr^{-1}$) and present scenario (106 ± 44 $km^2\ yr^{-1}$) (Table S4). This is likely related with the comparatively higher N inputs in intensively managed countries until roughly the 1990s (e.g., the Netherlands, France, Germany) relative to present conditions. More importantly, we did not find statistically significant differences between the three scenarios (p=0.73) (Table S5). Although this suggests that the trajectory of the predictors more directly influenced by the Nitrates Directive had a minor effect on hotspot size, we argue that it is not the case. Rather, reductions in N predictors must be spatially targeted according to environmental conditions (e.g., precipitation, soil texture) and legacy effect in groundwater.

Table S4. Quantification of annual hotspots in Europe for the two scenarios assuming N related predictors for historical conditions (average 1961-1990), present conditions (2000-2019).

Years	Historical conditions (1961-1990)	Present conditions (2000-2019)	Baseline
		thousand $km^2\ yr^{-1}$	
1992	101	96	111
1993	112	109	121
1994	166	160	175

1995	86	84	95
1996	70	69	81
1997	115	110	122
1998	76	73	86
1999	73	71	84
2000	90	85	94
2001	111	106	113
2002	78	77	80
2003	97	92	96
2004	68	66	67
2005	67	65	62
2006	74	71	66
2007	75	72	68
2008	79	78	73
2009	78	72	66
2010	78	75	71
2011	134	124	118
2012	108	102	95
2013	169	149	143
2014	167	148	142
2015	285	266	259
2016	161	136	126
2017	152	125	117
2018	184	154	142
2019	160	133	122

Table S5. Output from the ANOVA and Tukey test for Europe, which compared the two scenarios and baseline.

Anova					
Predictors	**Df**	**Sum Sq**	**Mean Sq**	**F value**	**Pr (>F)**
Periods	2	1296	648	0.317	0.729
Residuals	81	165723	2046		
Tukey test					
Variables	**Difference**	**Lower**	**Upper**	**p-value adjusted**	
Present-historical	-8.8	-37.6	20.1	0.75	
Baseline-historical	-7.8	-36.7	21.1	0.80	
Baseline-present	1.0	-27.9	30.0	0.99	

Countries

At the national level, our simulations showed marked variation in terms of the impact of the two scenarios (Fig. S20). The extent to which the baseline differs from the scenarios is influenced by the different trajectories relative to the historical and present averages (Fig. S21-22). On one hand, NO_3^- hotspots in groundwater did not show any significant response to the variation in the ND proxies (e.g., Greece, Bulgaria, Croatia, Romania): the scenarios and the baseline hotspot size were rather similar without any statistically significant differences (Table S6). We hypothesize that this may have been due to (i) the inability of European and national regulations to achieve a considerable reduction of e.g., N inputs, (ii) hotspots in these countries are mainly driven by external environmental conditions such as precipitation and soil texture. On the other hand, our simulations showed statistically significant differences in some countries (Tables S6-7). For instance, Portugal is a striking example as the hotspot size of the historical scenario was 2-3 times larger than the baseline or present scenario. This was also the case for Italy, Romania and Spain to a lesser extent. Conversely, the Netherlands showed marked trajectories with the baseline simulation yielding larger hotspot size than using the historical and present averages until roughly 2007. After 2007, there was a sharp decline relative to the scenarios.

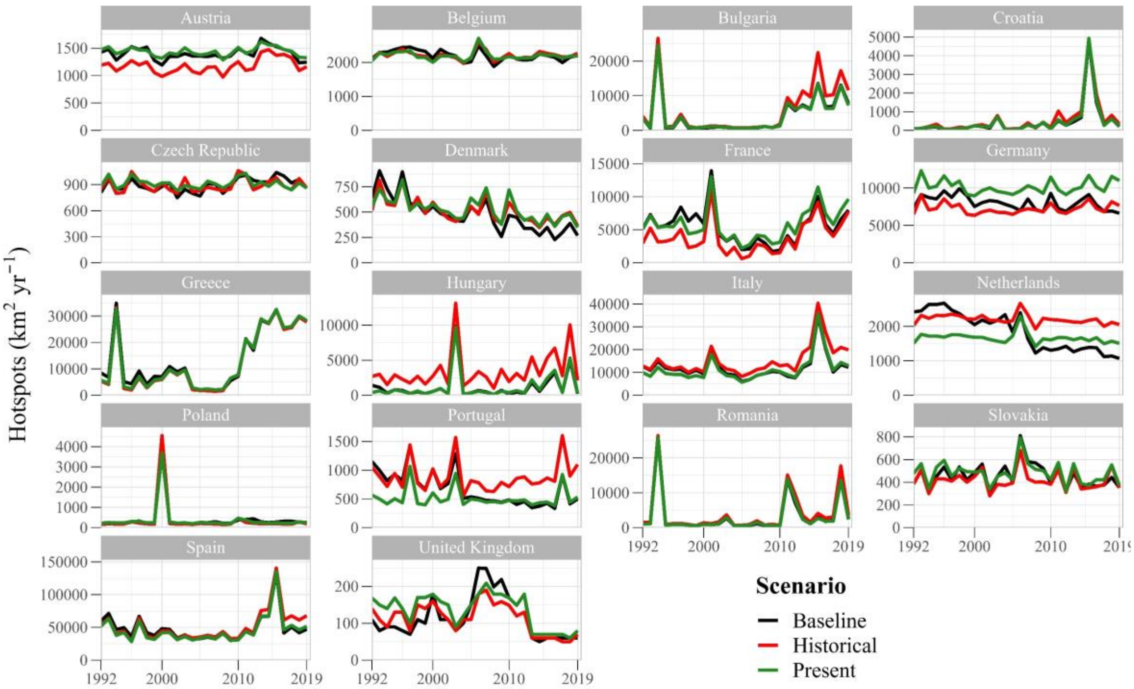

Figure S20. Hotspot size in different European countries for the period 1992-2019, accounting for the two scenarios: baseline (average 1961-1990), present (average 2000-2019) as well as baseline.

Table S6. P-values for the ANOVA between the two scenarios (historical, present) and baseline for each country.

Country	Pr (>F)
Spain	0.708
Italy	***0.048****
Greece	0.967
Bulgaria	0.584
France	***0.017****
Belgium	0.947
Hungary	***0.000*****
Poland	0.984
Croatia	0.906
Portugal	***0.000*****
Netherlands	***0.000*****
Czech Republic	0.568
Romania	0.886
Austria	***0.000*****
Slovakia	0.062
United Kingdom	0.202
Germany	***0.000*****
Denmark	0.450

Table S7. Tukey test between the two scenarios (historical, present) and baseline for each country where the null hypothesis was rejected (i.e., means between periods are equal).

Variables	Difference	Lower	Upper	p-value adjusted	Country
historical-baseline	2.7	-11.0	16.4	0.886	Spain
present-baseline	-2.1	-15.8	11.6	0.931	Spain
present-historical	-4.8	-18.5	8.9	0.686	Spain
historical-baseline	3.2	-0.7	7.0	0.127	Italy
present-baseline	-0.6	-4.5	3.3	0.926	Italy
present-historical	-3.8	-7.6	0.1	0.056	Italy
historical-baseline	-0.8	-7.9	6.4	0.964	Greece
present-baseline	-0.5	-7.6	6.7	0.986	Greece
present-historical	0.3	-6.9	7.5	0.995	Greece
historical-baseline	1.4	-2.4	5.3	0.646	Bulgaria
present-baseline	0.0	-3.9	3.8	1.000	Bulgaria
present-historical	-1.5	-5.3	2.4	0.636	Bulgaria
historical-baseline	***-1.6***	***-3.2***	***0.1***	***0.070****	***France***
present-baseline	0.4	-1.3	2.0	0.868	France

present-historical	***1.9***	***0.3***	***3.6***	***0.020****	***France***
historical-baseline	0.0	-0.1	0.1	0.999	Belgium
present-baseline	0.0	-0.1	0.1	0.961	Belgium
present-historical	0.0	-0.1	0.1	0.951	Belgium
historical-baseline	***2.3***	***0.8***	***3.7***	***0.001*****	***Hungary***
present-baseline	0.0	-1.4	1.4	1.000	Hungary
present-historical	***-2.3***	***-3.7***	***-0.8***	***0.001*****	***Hungary***
historical-baseline	0.0	-0.5	0.4	0.982	Poland
present-baseline	0.0	-0.5	0.5	0.994	Poland
present-historical	0.0	-0.5	0.5	0.996	Poland
historical-baseline	0.1	-0.5	0.7	0.907	Croatia
present-baseline	0.0	-0.6	0.6	0.995	Croatia
present-historical	-0.1	-0.7	0.5	0.942	Croatia
historical-baseline	***0.2***	***0.0***	***0.4***	***0.007****	***Portugal***
present-baseline	-0.2	-0.3	0.0	0.051	Portugal
present-historical	***-0.4***	***-0.5***	***-0.2***	***0.000*****	***Portugal***
historical-baseline	***0.3***	***0.1***	***0.6***	***0.001*****	***Netherlands***
present-baseline	-0.2	-0.4	0.0	0.100	Netherlands
present-historical	***-0.5***	***-0.8***	***-0.3***	***0.000*****	***Netherlands***
historical-baseline	0.0	0.0	0.0	0.989	Czech Republic
present-baseline	0.0	0.0	0.1	0.675	Czech Republic
present-historical	0.0	0.0	0.1	0.587	Czech Republic
historical-baseline	0.7	-3.1	4.4	0.909	Romania
present-baseline	0.0	-3.8	3.7	1.000	Romania
present-historical	-0.7	-4.4	3.0	0.901	Romania
historical-baseline	***-0.2***	***-0.3***	***-0.2***	***0.000*****	***Austria***
present-baseline	0.0	0.0	0.1	0.445	Austria
present-historical	***0.3***	***0.2***	***0.3***	***0.000*****	***Austria***
historical-baseline	0.0	-0.1	0.0	0.282	Slovakia
present-baseline	0.0	0.0	0.1	0.685	Slovakia
present-historical	0.1	0.0	0.1	0.053	Slovakia
historical-baseline	0.0	0.0	0.0	0.938	United Kingdom
present-baseline	0.0	0.0	0.0	0.363	United Kingdom
present-historical	0.0	0.0	0.1	0.209	United Kingdom
historical-baseline	***-0.8***	***-1.3***	***-0.2***	***0.003****	***Germany***
present-baseline	***2.1***	***1.6***	***2.7***	***0.000*****	***Germany***
present-historical	***2.9***	***2.3***	***3.4***	***0.000*****	***Germany***
historical-baseline	0.0	-0.1	0.1	0.692	Denmark

present-baseline	0.0	0.0	0.1	0.427	Denmark
present-historical	0.0	-0.1	0.1	0.903	Denmark
		*** p < 0.05 ** p<0.001**			

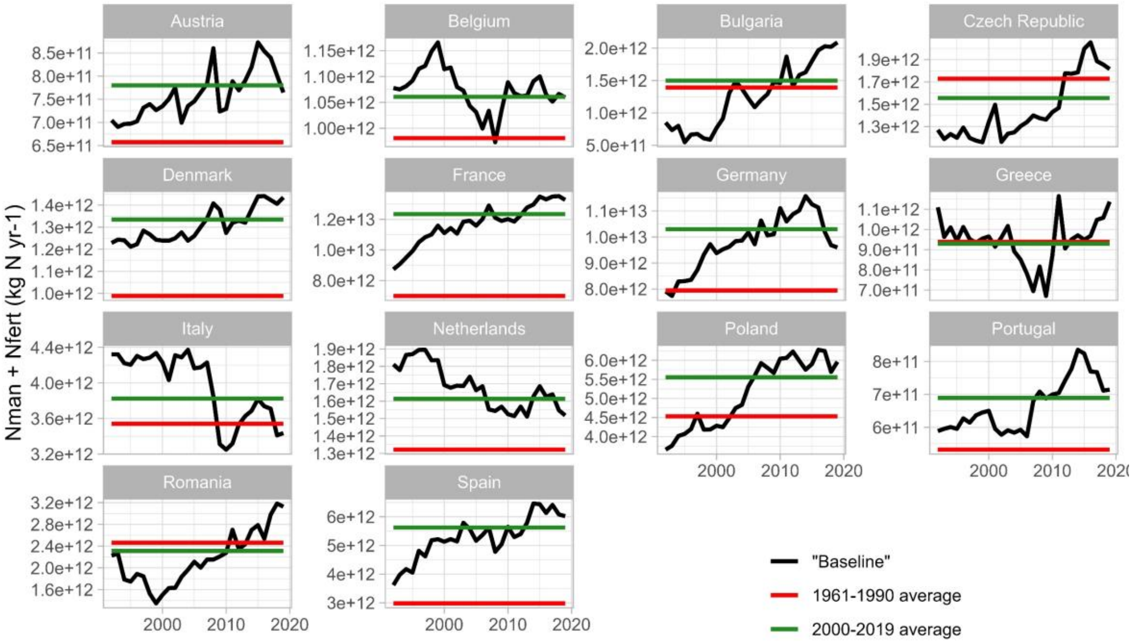

Figure S21. Different trajectories of the N inputs from manure and synthetic fertilisers in different European countries for the baseline, historical (1961-1990) and present (2000-2019) averages.

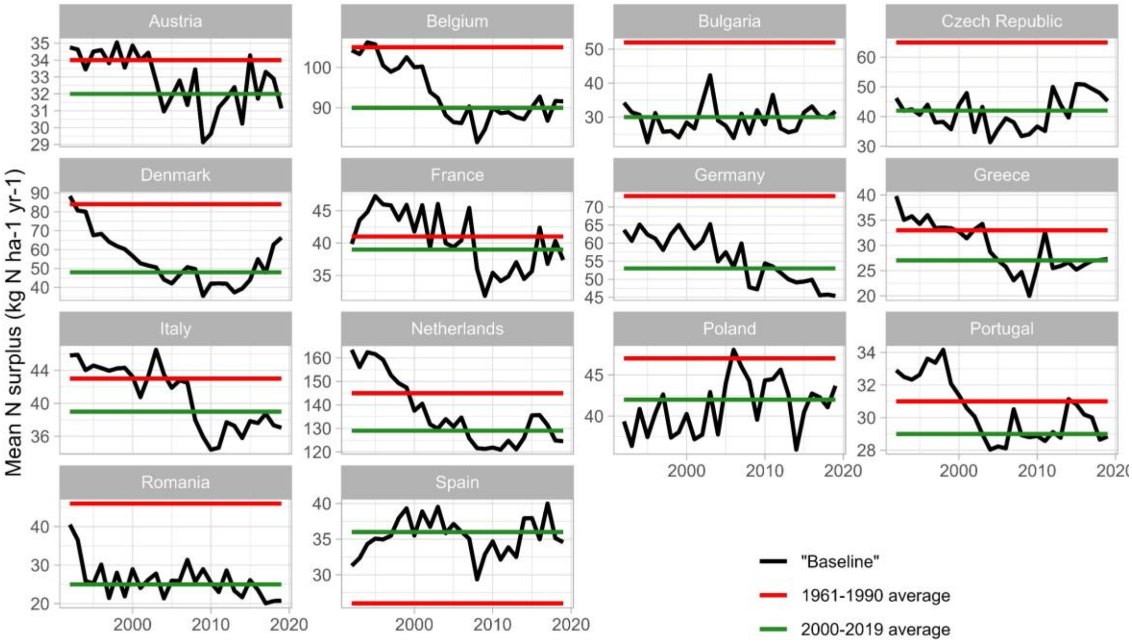

Figure S22. Different trajectories of the N surplus in different European countries for the baseline, historical (1961-1990) and present (2000-2019) averages.

8. Hotspot coverage in Europe

Table S8. Hotspot coverage of the different countries not using a whole territory approach.

Country	Average ±SD national hotspot area (thousand km^2)	Hotspot coverage (%)	NVZ area (thousand km^2) *	Fraction of the NVZ relative to total country area
Sweden	0	100	89.00	(20%)
United Kingdom	0.1 ± 0	96	96.00	(39%)
Estonia	≈ 0	100	3.00	(7%)
Slovakia	0.4 ± 0.1	93	24.00	(49%)
France	5 ± 3	90	292.00	(46%)
Czech Republic	1 ± 0	90	33.00	(42%)
Bulgaria	4 ± 6	82	25.00	(34%)
Hungary	1 ± 2	71	65.00	(70%)
Greece	14 ± 11	66	42.00	(32%)
Latvia	≈ 0	64	8.00	(13%)
Portugal	1 ± 0	57	4.00	(5%)
Spain	48 ± 20	47	122.00	(24%)
Italy	12 ± 6	39	46.00	(15%)
Croatia	1 ± 1	7	5.00	(9%)
Slovenia	≈ 0	0	0.03	(0.1%)

***** Number in parenthesis is the

9. Spatial variation of the require changes to meet water quality targets

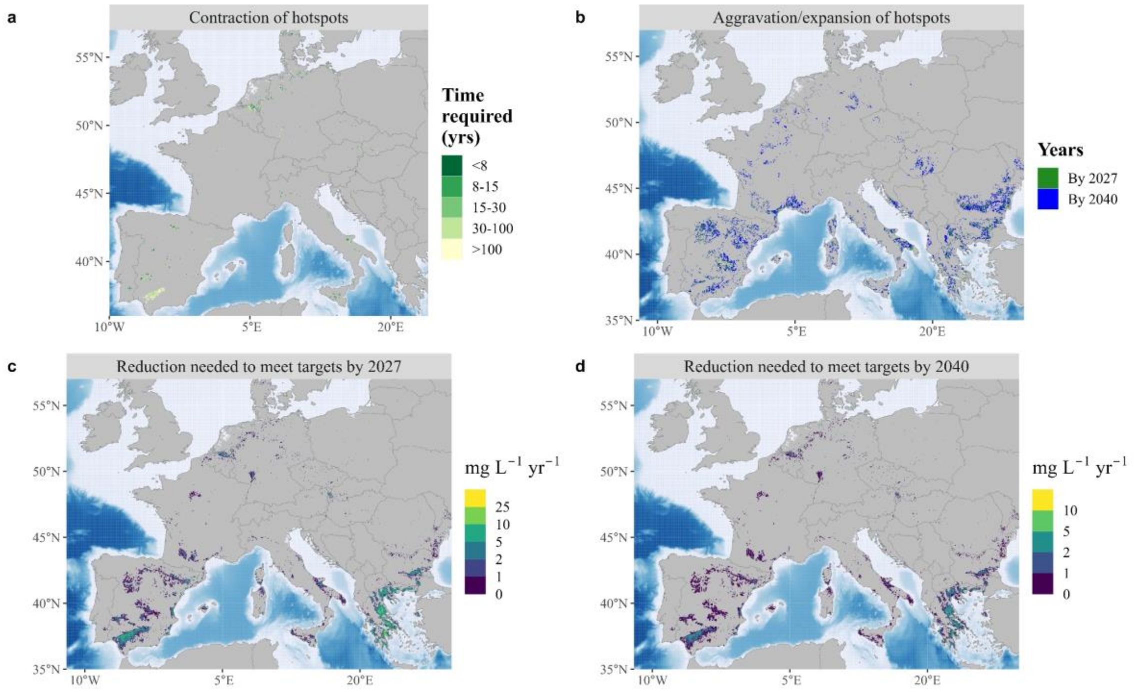

Figure S23. Time required until current hotspots (as of 2019) with declining trends can meet water quality standards (a). Expansion/aggravation of new hotspots by 2027 and 2040 (b). Reduction needed in current hotspots to meet water quality targets by 2027 (Water Framework Directive) and 2040 (long-term horizon) (c,d).

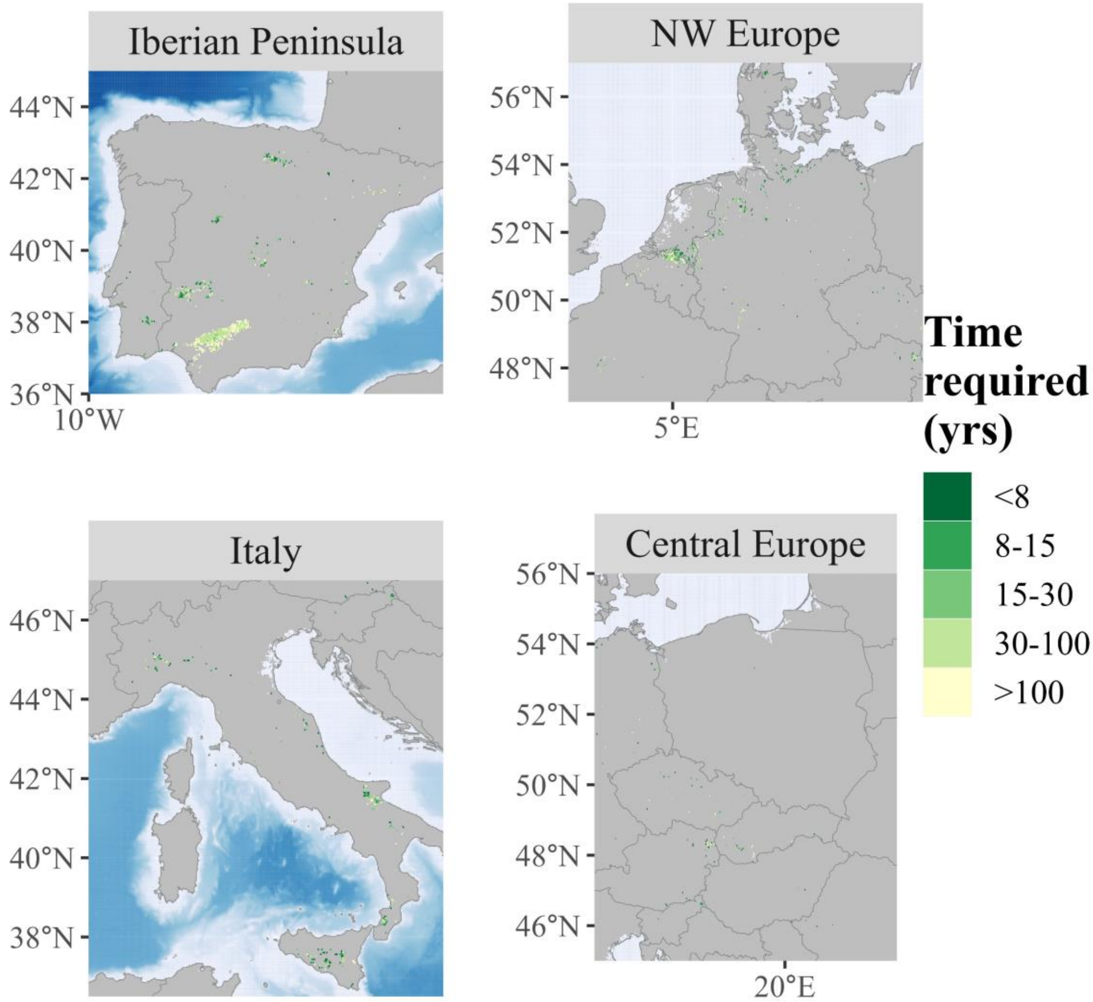

Figure S24. Zoom in on the spatial variation of the time required until current hotspots (as of 2019) with declining trends can meet water quality standards.