Characterizing Catchment-Scale Nitrogen Legacies and Constraining their Uncertainties

Fanny J. Sarrazin[1], Rohini Kumar[2], Nandita B. Basu[3], Andreas Musolff[2], Michael Weber[4], Kimberly Van Meter[5], and Sabine Attinger[4]

[1]Helmholtz-Centre for Environmental Research - UFZ
[2]Helmholtz Centre for Environmental Research - UFZ
[3]University of Waterloo
[4]Helmholtz-Centre for Environmental Research -UFZ
[5]Pennsylvania State University

November 21, 2022

Abstract

Improving nitrogen (N) status in European water bodies is a pressing issue. N levels depend not only on current but also past N inputs to the landscape, that have accumulated through time in legacy stores (e.g. soil, groundwater). Catchment-scale N models, that are commonly used to investigate in-stream N levels, rarely examine the magnitude and dynamics of legacy components. This study aims to gain a better understanding of the long-term fate of the N inputs and its uncertainties, using a legacy-driven N model (ELEMeNT) in Germany's largest national river basin (Weser; 38,450 km2) over the period 1960-2015. We estimate the nine model parameters based on a progressive constraining strategy, to assess the value of different observational datasets. We demonstrate that beyond in-stream N loading, soil N content and in-stream N concentration allow to reduce the equifinality in model parameterizations. We find that more than 50% of the N surplus denitrifies (1480-2210 kg ha-1) and the stream export amounts to around 18% (410-640 kg ha-1), leaving behind as much as around 230-780 kg ha-1 of N in the (soil) source zone and 10-105 kg ha-1 in the subsurface. A sensitivity analysis reveals the importance of different factors affecting the residual uncertainties in simulated N legacies, namely hydrologic travel time, denitrification rates, a coefficient characterising the protection of organic N in source zone and N surplus input. Our study calls for proper consideration of uncertainties in N legacy characterization, and discusses possible avenues to further reduce the equifinality in water quality modelling.

Characterizing Catchment-Scale Nitrogen Legacies and Constraining their Uncertainties

F. J. Sarrazin[1], R. Kumar[1], N. B. Basu[2,3,4], A. Musolff[5], M. Weber[1], K. Van Meter[6,7], S. Attinger[1,8]

[1]Department of Computational Hydrosystems, Helmholtz-Centre for Environmental Research - UFZ, Leipzig, Germany
[2]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada
[3]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada
[4]Water Institute, University of Waterloo, Waterloo, ON, Canada
[5]Department of Hydrogeology, Helmholtz-Centre for Environmental Research - UFZ, Leipzig, Germany
[6]Department of Geography, Pennsylvania State University, PA, USA
[7]Earth and Environmental Systems Institute, Pennsylvania State University, PA, USA
[8]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Key Points:

- We use a parsimonious model to examine the long-term (1960–2015) fate of nitrogen inputs to the landscape including the legacy buildup
- More than 50% of the nitrogen surplus denitrifies and around 491 $kg\ ha^{-1}$ accumulates in legacy stores in Germany's largest river basin
- Hydrologic travel time and denitrification rates largely contribute to the residual uncertainty in the simulated nitrogen legacies

Corresponding author: Fanny Sarrazin, `fanny.sarrazin@ufz.de`

Corresponding author: Rohini Kumar, `rohini.kumar@ufz.de`

Abstract

Improving nitrogen (N) status in European water bodies is a pressing issue. N levels depend not only on current but also past N inputs to the landscape, that have accumulated through time in legacy stores (e.g. soil, groundwater). Catchment-scale N models, that are commonly used to investigate in-stream N levels, rarely examine the magnitude and dynamics of legacy components. This study aims to gain a better understanding of the long-term fate of the N inputs and its uncertainties, using a legacy-driven N model (ELEMeNT) in Germany's largest national river basin (Weser; 38,450 km^2) over the period 1960–2015. We estimate the nine model parameters based on a progressive constraining strategy, to assess the value of different observational datasets. We demonstrate that beyond in-stream N loading, soil N content and in-stream N concentration allow to reduce the equifinality in model parameterizations. We find that more than 50% of the N surplus denitrifies (1480–2210 $kg\ ha^{-1}$) and the stream export amounts to around 18% (410–640 $kg\ ha^{-1}$), leaving behind as much as around 230–780 $kg\ ha^{-1}$ of N in the (soil) source zone and 10–105 $kg\ ha^{-1}$ in the subsurface. A sensitivity analysis reveals the importance of different factors affecting the residual uncertainties in simulated N legacies, namely hydrologic travel time, denitrification rates, a coefficient characterising the protection of organic N in source zone and N surplus input. Our study calls for proper consideration of uncertainties in N legacy characterization, and discusses possible avenues to further reduce the equifinality in water quality modelling.

Plain Language Summary

Lowering nitrogen (N) amounts in European surface waters is a pressing issue. N levels largely result from fertilizer application in agricultural areas, and deposition of atmospheric N coming from fossil fuel combustion. These N inputs to the landscape can accumulate below the ground surface in so-called legacy stores (including the soil and aquifer), from which they can be released progressively through time. Therefore, N levels depend not only on the recent N inputs, but also on their history. Our modelling study aims to improve our understanding of the long-term fate of the N inputs and its uncertainties in Germany's largest national river basin (Weser) over the period 1960–2015. It suggests that more than 50% of the N inputs to land is lost to the atmosphere (denitrification, 1480–2210 $kg\ ha^{-1}$) and the stream export amounts to around 18% (410–640 $kg\ ha^{-1}$), leaving behind as much as around 16% (264–820 $kg\ ha^{-1}$) in the landscape (legacy). However, the uncertainties in these estimates remain large, partly due to a lack of observational data on internal (legacy)

components and uncertainties in N inputs. Overall, our study calls for proper consideration of uncertainties in N legacy characterization, and discusses possible avenues to further reduce them.

1 Introduction

Since the beginning of the 20th century, nitrogen (N) inputs to soils have increased tremendously worldwide, due to application of synthetic N fertilizers and manure in agricultural areas (Lu & Tian, 2017; B. Zhang et al., 2017), and due to deposition of atmospheric N, coming mostly from fossil fuel combustion (Holland et al., 1999). These large N inputs to soils have not been balanced by N removal from soils through incorporation into plant biomass and crop harvest, resulting in large soil N surpluses in many places across the world, e.g. in the US (Byrnes et al., 2020), in Europe (Erisman et al., 2011) and in China (X. Wang et al., 2014). Furthermore, wastewater can be released to the environment with or without treatment, which constitutes a point source of N (United Nations, 2017). In Europe, the efficiency of wastewater treatment plants has largely improved since 1990 (European Commission, 2020). Nevertheless water bodies still receive large point N loads, and for many countries, further enhancement of treatment facilities are prescribed (European Commission, 2020; Svirejeva-Hopkins et al., 2011).

This excess of N in the environment has substantial consequences for both the hydrosphere and the atmosphere (Robertson & Groffman, 2015). In fact, soluble N compounds, and more specifically nitrate, are polluting drinking water, which threatens human health (WHO, 2016). They are also causing eutrophication of receiving water bodies, such as European seas (EEA, 2019b) and coastal areas in the US (Scavia & Bricker, 2006). In addition, denitrification in the terrestrial system, which microbially reduces nitrate, can release large amount of nitrous oxide (N_2O) (Tian et al., 2018). The latter is an important greenhouse gas almost 300 times as potent as CO_2 (Myhre et al., 2013), and contributes to the depletion of stratospheric ozone (Portmann et al., 2012). Therefore, reducing nitrate levels through improved land management strategies is a pressing issue, in particular in Europe where countries such as France, Germany and Greece have been recently fined by the European Court of Justice for exceeding the regulatory limits for nitrate in receiving water bodies (Damania et al., 2019).

Sporadic evidence shows that N inputs to the soil can not only reach surface water bodies or be lost to the atmosphere via denitrification, but they can also accumulate in the landscape as legacy N (Chen et al., 2018; Van Meter & Basu, 2015). A few studies including long-term measurements of soil N content in the Mississippi River Basin (Van Meter et al., 2016), fertilization experiments in the UK and in France (Jenkinson, 1991; Sebilo et al., 2013) and mass balance assessments (Smil, 1999; Worrall et al., 2015) suggest that part of the N surpluses can build up in the root zone of agricultural soils in organic forms (biogeochemical legacy). Furthermore, increasing nitrate concentration trends are detected in groundwater in some locations in the US and the UK (Puckett et al., 2011; Stuart et al., 2007). The pervasive issue of nitrate pollution in groundwater, in particular in Europe (Sundermann et al., 2020; EEA, 2018), and the results of modelling studies focusing on N accumulation in the vadose zone (Ascott et al., 2017; L. Wang et al., 2012) points to a widespread buildup of N in dissolved inorganic forms (hydrologic legacy). Biogeochemical and hydrologic legacies can impact the future water quality status of water bodies. They can induce a delay or "time lag" between changes in land management practices and the water quality response (Grimvall et al., 2000; Vero et al., 2018). The implications of the two legacy forms for future water quality could be very different because hydrologic legacy corresponds to the accumulation of reactive and mobile N, and biogeochemical legacy to the accumulation of much more stable (organic) N compounds. Therefore, a characterization of both types of N legacies at catchment scale is crucial to inform the design of effective water quality measures.

However, our understanding of the magnitude of these legacies and their associated timescale remains limited because of a lack of observational data and the uncertainties associated with N legacy modelling. Direct observational data of N content in soil and groundwater are often sparse in time and space. This makes it difficult to capture the temporal changes and spatial variability in N content and to determine the integrated legacy behaviour at catchment scale. In particular, large-scale datasets of soil N content are often available for one year (e.g. Ballabio et al., 2019). Regarding groundwater N concentration measurements, in Europe, low-frequency (typically annual) measurements and the fact that they are mostly available for the recent years impedes the analysis of the long-term dynamics. Additionally, assessing the spatial distribution of groundwater concentration from point measurements involves large uncertainties (see e.g. Figure 6 in Knoll et al., 2020). Therefore, mechanistic models should be used to quantify N legacies, to complement the information

provided by currently available observation data of N content in soil and groundwater that are insufficient.

In this respect, some past modelling studies investigate the time lag between the trajectory of N inputs to the terrestrial system and the in-stream N levels using empirical approaches, e.g. the studies of Chen et al. (2014), Dupas et al. (2020), Ehrhardt et al. (2021) and Van Meter and Basu (2017) (further details are reported in the review of Chen et al., 2018). Yet, such modelling approaches are typically based on lumped transfer functions that relate the N inputs to the N stream export at catchment scale and they do not explicitly disentangle the role of biogeochemical and hydrologic legacies. Only few studies, all of them focusing on North America, explicitly consider and examine both types of N legacies at catchment scale using mechanistic models, namely a modified version of the Soil Water Assessment Tool model (SWAT-LAG, Ilampooranan et al., 2019), the NOAA's Geophysical Fluid Dynamics Laboratory Land Model (LM3-TAN, Lee et al., 2016), and the more parsimonious Exploration of Long-tErM Nutrient Trajectories model (ELEMeNT, Chang et al., 2021; J. Liu et al., 2021; Van Meter et al., 2017, 2018).

Importantly, the investigation of N legacies through modelling approaches is fraught with large uncertainties. First, the N input data have large uncertainties because their construction involves numerous uncertain factors, as reported in Byrnes et al. (2020), Häußermann et al. (2020), Hong et al. (2012), Poisvert et al. (2017), and Häußermann et al. (2020) for diffuse N sources and in Grizzetti et al. (2008), Morée et al. (2013), and Van Meter et al. (2017) for point sources (wastewater). Notably, although the application of mineral fertilizer is a key N input to agricultural soils, data may only be available at the national level for some countries such as Germany, and spatial disaggregation strategies have to be developed to estimate fertilizer application at finer spatial resolutions (Häußermann et al., 2020). Second, the modelling of N legacies suffers from a lack of process understanding, e.g. regarding the immobilization and accumulation of N into organic matter in soils (see e.g. Bingham & Cotrufo, 2016; Yansheng et al., 2020). Third, the parameters of mechanistic N models are generally estimated through calibration (Moriasi, Zeckoski, et al., 2015), because they are often conceptual parameters that cannot be directly related to measurable quantities. Parameter values are typically constrained using in-stream measurements only (Moriasi, Zeckoski, et al., 2015), since observational data of internal model variables, such as N content in soil and groundwater, are generally lacking. It is well known that different parameter sets can fit the in-stream data equally well, as discussed in previous water quality

studies, e.g. Ford et al. (2017), Husic et al. (2019), Rankinen et al. (2006), and Wade et al. (2008). Because of this issue of equifinality (Beven, 2006), it may be possible to identify a range of plausible values for the model internal fluxes and states corresponding to the different plausible parameter sets. Hence, simulated N legacies (internal model states) might be poorly characterized if in-stream information only is used to constrain model parameterizations. All of the discussed uncertainties in N legacy modelling are exacerbated by the fact that the N accumulation in the landscape needs to be modelled over long timescales (decades to centuries) to understand the contemporary and future water quality status. In fact, long time series of input data are fraught with uncertainties and long records of observations of model outputs, such as in-stream N concentrations, are rarely available.

Our review of the literature suggests that 1) we lack understanding of the magnitude and timescale of N legacies at catchment scale in light of the associated uncertainties and 2) it remains unclear whether and to what extent the in-stream information only, as typically used to calibrate catchment-scale N water quality models, is able to constrain the simulated N legacies and which additional information would mostly help to reduce uncertainty. In this study, we address these gaps by investigating the long-term fate of the N inputs to the landscape and its uncertainties. In particular, we analyze the uncertainties in simulated N biogeochemical and hydrologic legacies due to the uncertainties in the model parameters and the input data, and we determine the value of different types of (observational) data to constrain the modelling results.

To this end, first, we introduce a multicriteria approach based on soft rules to constrain the model parameters, which allows assessment of parameter uncertainty and of the value of the different observational data available (in-stream N loading and concentration data, and soil N data). Secondly, we perform a sensitivity analysis to determine the factors responsible for the (residual) uncertainty in the simulated N legacies to prioritize future efforts for uncertainty reduction and model improvement. We apply the ELEMeNT model, which is a parsimonious N model that explicitly accounts for both biogeochemical and hydrologic legacies (Van Meter et al., 2017). While past modelling studies focus on North America (Chang et al., 2021; Ilampooranan et al., 2019; Lee et al., 2016; J. Liu et al., 2021; Van Meter et al., 2017, 2018), here we extend the analyses of N legacies to the European context. We examine the N legacy behaviour over the last six (1960–2015) decades in the Weser river basin (WRB), which is Germany's largest national river basin. Through the application of the soft rules and the sensitivity analysis, we infer the dominant factors

affecting the uncertainty in simulated N legacies and discuss the implications for future modelling studies, data monitoring, and water and land management strategies.

2 Model description

The ELEMeNT model (Van Meter et al., 2017) simulates the fate of the diffuse N sources (soil N surplus) and point sources at annual timescale and it computes the annual in-stream nitrate-N loading and concentration at the catchment outlet. The model assumes that dissolved N is present in the form of nitrate (NO_3). ELEMeNT conceptualizes the N processes in two terrestrial compartments, a shallower compartment referred to as "source zone", which represents the soil root zone (assumed to have a depth of 1 m), and a deeper compartment referred to as "subsurface zone", which includes both the unsaturated zone below the source zone and the groundwater. The land use categories considered are cropland, agricultural permanent grassland, and non-agricultural land (see Section 3.2.1). ELEMeNT requires as input the annual time series of the different land use fractions, the N surplus for each land use category, the stream discharge at the catchment outlet, and the N point sources. We revise the formulation of the subsurface submodel and add an in-stream submodel as detailed in Sections 2.2 and 2.3. The model version used in this study counts nine calibration parameters (Table 1).

Table 1. Description of the ELEMeNT Model Parameters, Ranges Used in this Study, and References for the Determination of the Ranges

Parameter	Description	Unit	Lower value	Upper value	References for the values
$M_{s_{org}}^{prist}$	Source zone organic N stock under pristine conditions	$(kg\ ha^{-1})$	10^4	3.5×10^4	The range includes the estimates of soil N content in 2009 from LUCAS dataset (Ballabio et al., 2016, 2019)
h_c	N protection coefficient for cultivated land	(-)	0.1	0.5	Van Meter et al. (2017, Table S2.1); Chang et al. (2021, Table S3)
h_{nc}	N protection coefficient for non-cultivated land	(-)	0.25	0.75	Van Meter et al. (2017, Table S2.1); Chang et al. (2021, Table S3)
k_a	Mineralization rate constant for organic active N store	(yr^{-1})	0.05	0.75	The range includes the values in Van Meter et al. (2017, Table S2.1)
V_s [a]	Mean annual water content in the source zone	(mm)	100	500	Van Meter et al. (2017, Table S2.1)
λ_s	Denitrification rate constant in the source zone	(yr^{-1})	0.1	1	Van Meter et al. (2017, Table S2.1)
λ_{sub}	Denitrification rate constant in the subsurface zone	(yr^{-1})	0.01	0.3	Van Meter et al. (2017, Table S2.1); Puckett et al. (2011); Heidecke et al. (2015, Sect. 3.2.3.2)
μ_{sub}	Mean travel time in the subsurface	(yr)	2	50	Van Meter et al. (2017, Table S2.1); Koeniger et al. (2008, Table 2)
R [b]	Fraction of in-stream N removal	(-)	0.01	0.3	Howarth et al. (1996, Table 9); Mulholland et al. (2008, Figure 4.c); X. Yang et al. (2018); Grizzetti et al. (2008, Table 6)

[a] V_s integrates the parameters for the porosity, saturation and depth of the source zone reported in Van Meter et al. (2017).

[b] R is a new parameter for in-stream processes, introduced in this study.

2.1 Source Zone Submodel

Following Van Meter et al. (2017), the model assesses the dynamics of the organic N pool, accounting for the biogeochemical legacy, and of the inorganic N pool. The organic N pool is divided into two different stores: (1) a protected N store with a slow turnover, that accounts for the physical stabilization of N, and (2) an active organic N store, that consists of organic N which is more labile and prone to mineralization. The organic N stores receive the total N surplus. Therefore mineralization in the model (the simulated N mass flux from the organic to the inorganic pool) represents the effective N mass flux to the inorganic N pool. This flux results from the combined effect of the transformation of organic N into mineral forms and the immobilization of mineral N inputs into the organic matter. Immobilization is understood to be an important process in both agricultural soils (Blankenau et al., 2001; Haag & Kaupenjohann, 2001; X. Liu et al., 2016; Macdonald et al., 1989; Sebilo et al., 2013; Yansheng et al., 2020) and forested soils (Castellano et al., 2012; Lewis & Kaye, 2012; Morier et al., 2008; Spoelstra et al., 2001). However, these process interactions are yet not well understood (Bingham & Cotrufo, 2016; Yansheng et al., 2020) and therefore difficult to represent explicitly in a mechanistic model. Since mineralization is parameterized as a first order process, part of the N surplus is quickly transferred to the inorganic N store and can therefore be readily available for leaching. A protection coefficient determines the fraction of the N surplus that is added to the protected store. Transfer of N mass from the protected to the active N stores occurs in case of land use conversion from uncultivated land to cropland. After mineralization occurs, the inorganic N in the source zone can either denitrify or leach to the subsurface.

2.2 Subsurface Submodel

In the subsurface, the transport of dissolved inorganic N is represented using a travel time distribution to account for hydrologic legacy, while N can denitrify following a first order process (Van Meter et al., 2017). In this study, we revise the formulation for the subsurface compartment. Compared to the previous model formulation (Van Meter et al., 2017), the travel time distribution function is an explicit function of time t (Benettin et al., 2015; Botter et al., 2010; Queloz et al., 2015). The N-NO_3 export from the subsurface to the stream $J_{stream_{sub}}$ *(t)* ($kg\ ha^{-1}\ yr^{-1}$) at time t is therefore written as:

$$J_{stream_{sub}}(t) = \int_0^{+\infty} J_{sub_s}(t-T)p(T,t-T)e^{-\lambda_{sub}T}dT \quad (1)$$

where T (yr) is the travel time, J_{sub_s} $(kg\ ha^{-1}\ yr^{-1})$ is the N-NO_3 mass leaching from the source zone to the subsurface, $p(T,t)$ (-) is the travel time distribution function and λ_{sub} (yr^{-1}) is the rate constant of denitrification in the subsurface. As in Van Meter et al. (2017), we assume complete mixing (or random sampling) in the subsurface compartment and we adopt an exponential distribution for the travel time (Equation 41 in Botter et al., 2010). The mean value of the distribution $\mu'(t)$ (yr) is given as:

$$p(T, t-T) = \frac{1}{\mu'(t)} e^{-\int_{t-T}^{t} \frac{1}{\mu'(x)} dx} \tag{2}$$

with

$$\mu'(t) = \frac{\overline{Q_{out}}\mu_{sub}}{Q_{out}(t)} \tag{3}$$

where $Q_{out}(t)$ $(mm\ yr^{-1})$ is the discharge at the catchment outlet, $\overline{Q_{out}}$ $(mm\ yr^{-1})$ is the arithmetic mean of the discharge, and μ_{sub} (yr) is the harmonic mean of $\mu'(t)$. To account for the fact that the mean discharge may not be stationary, we compute $\overline{Q_{out}}$ at each time step as the 30-year backward moving average of the discharge. We refer to supplementary Sections S1-S3 for further details on the mathematical derivation and numerical integration of the new equations introduced in this study.

2.3 In-stream Submodel

In-stream N removal, which comprises the processes of denitrification and biotic assimilation described e.g. in Basu et al. (2011) and Dehaspe et al. (2021), was implicitly accounted for in Van Meter et al. (2017). The removal of N from point sources was lumped with the removal due to wastewater treatment, and the removal of N from diffuse sources was lumped with the denitrification in the terrestrial compartments. In this study, we need to represent in-stream removal explicitly, since our point N sources input, described in Section 3.2.4, already includes the N removal through wastewater treatment. The in-stream NO_3-N load $J_{out}(t)$ $(kg\ ha^{-1}\ yr^{-1})$ is computed as:

$$J_{out}(t) = (J_{stream_{sub}}(t) + J_{ps}(t))R \tag{4}$$

where R (-) is the annual fraction of in-stream N removal and $J_{ps}(t)$ $(kg\ ha^{-1}\ yr^{-1})$ is the N loading from point sources. For our application at catchment scale and annual timescale, we thereby assume that in-stream N removal can be represented by a first order process, as documented e.g. in Basu et al. (2011).

3 Study Catchments and Data

3.1 Description of the Study Catchments

We apply the ELEMeNT model to the Weser river basin (WRB) in Germany, focusing on the region located upstream of the Hemelingen station (see Figure 1a, b). This covers an area of around 38,450 km^2, which corresponds to almost 11% of the total area of Germany. The Weser river largely contributes to the total N load discharging into the North Sea, where eutrophication is a major issue (Arle et al., 2017). In the WRB, a priority goal set by the WRB Commission (FGG Weser, 2020) is to reduce N inputs to the landscape to achieve in-stream N concentrations below the regulatory threshold of 2.8 $mg\ L^{-1}$ (OGewV, 2016). For our analyses, we selected eight stations, including Hemelingen, that have a long record (between 26 and 37 years) of 14-days average in-stream nitrate measurements, that were constructed by mixing daily samples. These stations are situated on the Weser, Werra, Fulda and Aller rivers and their location is reported in Figure 1b.

The WRB is characterized by a humid temperate climate with an annual mean precipitation around 780 $mm\ yr^{-1}$ (spatial range: 600–1100 $mm\ yr^{-1}$) and a mean aridity index (ratio of potential evapotranspiration to precipitation) around 0.9 (S. Yang et al., 2019; Zink et al., 2017). The average annual discharge at Hemelingen over the period 1950–2015 amounts to 268 $mm\ yr^{-1}$. Agriculture is the dominant land use type and constitutes 45% of the catchment area in 2015 (35% cropland and 10% agricultural permanent grassland). Other important land uses are forested land and other vegetated land (including natural grassland and urban green areas), which cover 34% and 19% of the catchment area in 2015, respectively. Spatially and temporally averaged annual N surplus is estimated to be equal to 48.6 $kg\ ha^{-1}\ yr^{-1}$ over the period 2011–2015, and N surplus takes a higher value over agricultural areas (70.9 $kg\ ha^{-1}\ yr^{-1}$) compared to non-agricultural areas (30.3 $kg\ ha^{-1}\ yr^{-1}$). On average in-stream N-NO_3 concentration (C_{out}) value at the catchment outlet is equal to 3.7 $mg\ L^{-1}$ over the period 2011–2015, which is above the target value of 2.8 $mg\ L^{-1}$ (OGewV, 2016).

The eight nested subcatchments analyzed in this study, present some moderate differences in their characteristics, as indicated in Figure 1c-j. In particular, the percentage of agricultural areas ranges from 38% to 48%, with lower values in the southern (upstream) part, which is situated in a mountainous area, compared to the northern (downstream) part, which lies in the German Northern Plain. We notice a sharp decrease in N surplus in 1990

for the Letzter Heller subcatchment (Figure 1j). This can be explained by the fact that a large area of the subcatchment is situated within the state of Thuringia, which was part of the former German Democratic Republic, and where agricultural activities were profoundly disrupted following the German reunification in 1990. The in-stream concentration C_{out} ranges from 3.0 $mg\ L^{-1}$ at the Wahnhausen station (Figure 1i), which is located in the upstream part, to 4.1 $mg\ L^{-1}$ at the downstream Drakenburg and Porta stations (Figure 1d, f).

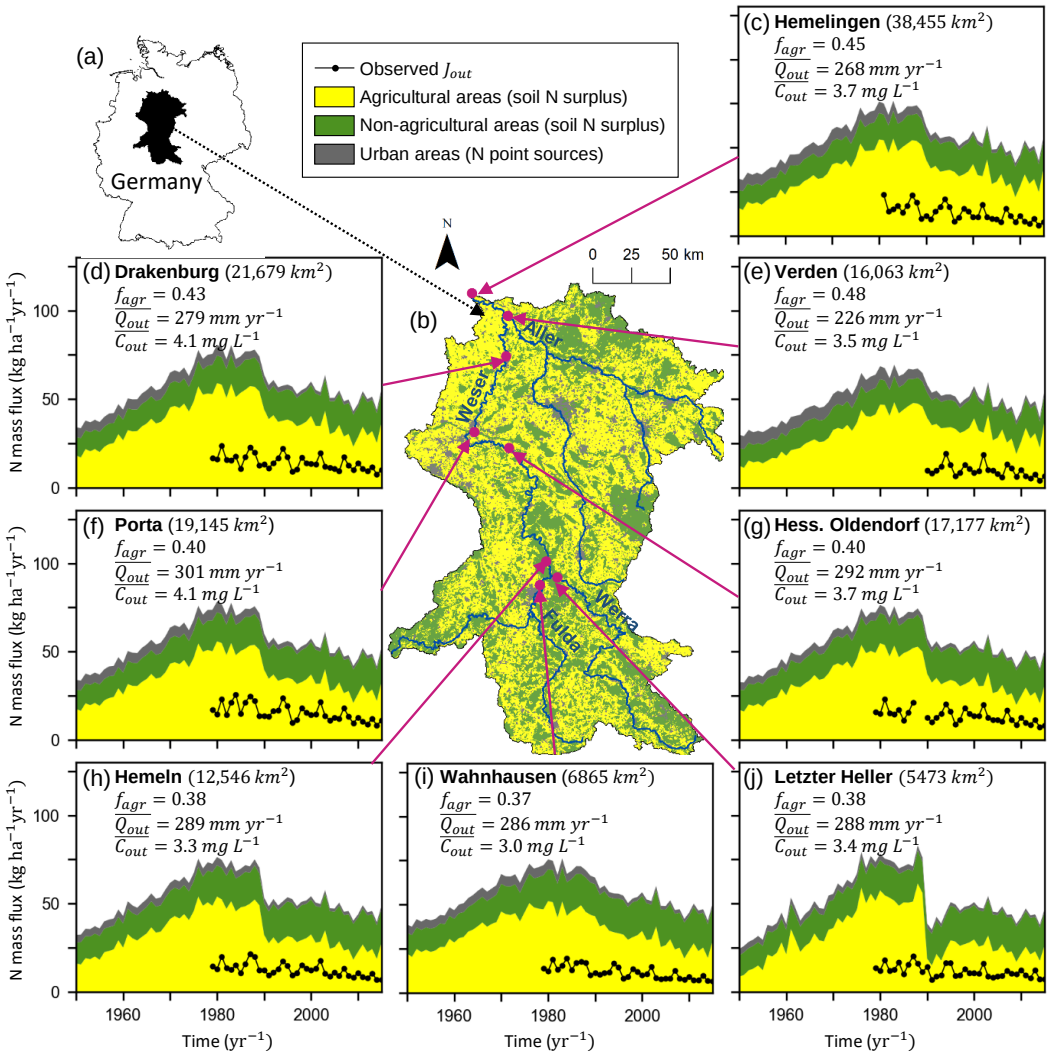

Figure 1. The eight subcatchments of the WRB selected for the analyses: (a) location of the WRB in Germany, (b) land use of the WRB and outlet of the subcatchments analyzed, and (c-j) N input (N surplus and N point sources), in-stream observations (N-NO_3 loading, J_{out}) time series for the period 1950–2015, and catchment properties namely catchment area, fraction of agricultural area in 2015 (f_{agr}), average annual stream discharge over the period 1950–2015 ($\overline{Q_{out}}$), and average observed in-stream N-NO_3 concentration over the period 2011–2015 ($\overline{C_{out}}$). Source of the land use data in panel (b): Corine Land Cover data for year 2012 (EEA, 2019a). Other datasets used are described in Section 3.2.

3.2 Data Description and Processing

In this section, we present the datasets adopted to run ELEMeNT and the datasets of N content observations used to assess the performance of the model and estimate its parameters (as explained in Section 4).

3.2.1 Land Use Data

We construct the 1800–2015 trajectories of the catchment-scale fractions of the three land use categories required by ELEMeNT, namely cropland, agricultural permanent grassland and non-agricultural land. The latter land cover includes forest, other vegetated land (such as natural grassland, green urban areas), built-up areas and non-vegetated land. We combine data from the gridded Corine Land Cover dataset (CLC; EEA, 2019a), the gridded History Database of the Global Environment dataset (HYDE; Klein Goldewijk et al., 2011, 2017) and census data on agricultural areas available from the Federal Statistical Office (Statistisches Bundesamt, 2021) and from the yearly statistical books for Germany (Digizeitschriften, 2021). Supplementary Section S4 provides details on the methodology used to construct the land use data. The trajectories of the land use fractions for the different subcatchments are presented in supplementary Figure S1.

3.2.2 N Surplus for Agricultural Areas

We adopt the N surplus dataset for agricultural areas of Häußermann et al. (2020), which is available for the period 1995–2015 at the county level, and the dataset of Behrendt et al. (2003), which is available for the period 1950–1998 at the state level, and which builds on the studies of Bach and Frede (1998) and Behrendt et al. (2000). The components included in the two datasets are the N content in the input of mineral fertilizer, manure, other organic fertilizer such as sewage sludge, seeds and planting material (for the county level dataset only), atmospheric deposition, biological fixation by legumes, as well as the N content in harvested crops. We refer to Häußermann et al. (2020) and Behrendt et al. (2003) for further methodological details. We harmonize the two N surplus datasets similar to Ehrhardt et al. (2021), to construct the N surplus trajectories at county level for the period 1950–2015 (see Supplementary Section S5 for further details). The resulting trajectories of the N surplus for agricultural areas are shown in supplementary Figure S2.

3.2.3 N Surplus for Non-agricultural Areas

Following Van Meter et al. (2017), for non-agricultural areas, we consider two components for the N surplus, namely atmospheric N deposition and biological N fixation, and we neglect any net accumulation of N in the vegetation. We quantify atmospheric N deposition using the dataset produced from Community Atmosphere Model with Chemistry (CAM-chem, Lamarque et al., 2012) simulations, as part of the National Center for Atmospheric Research (NCAR) Chemistry-Climate Model Initiative (CCMI, Tilmes et al., 2016) N deposition dataset. This product is part of the input datasets for the Model Intercomparison Projects (input4MIPS) and it is a forcing dataset for the Coupled Model Intercomparison Project phase 6 (CMIP6, Eyring et al., 2016). The data are provided over a $1.9° \times 2.5°$ grid for the period 1850–2014. We estimate biological N fixation using the mean annual rate reported by Cleveland et al. (1999) for natural temperate forest ($16\ kg\ ha^{-1}\ yr^{-1}$) and for natural grassland ($2.7\ kg\ ha^{-1}\ yr^{-1}$). We use the latter for our land use category "other vegetated land" defined in Section 3.2.1. The trajectories of atmospheric N deposition and biological N fixation are shown in supplementary Figure S3.

3.2.4 N Point Sources

For N point sources, we use observations of N loading, available for 802 wastewater treatment plants (WWTPs) and one year in the period 2012–2016 depending on the plants (S. Yang et al., 2019; Büttner, 2020). Data for the larger WWTPs (with population equivalent over 2,000) come from the Environmental Agency database (EEA, 2015) and correspond to the year 2012, while data for the smaller WWTPs (with population equivalent under 2,000) come from the authorities of the federal German states and correspond to the year 2015 or 2016. For the past years, we estimate N point sources from wastewater from the methodology proposed by Morée et al. (2013). We utilize data on population count (HYDE dataset), protein supply (FAO, 1951, 2021a, 2021b), and population connection to sewer and WWTPs (Seeger, 1999; Eurostat, 2016, 2021). We combine these data and create an ensemble of historical N loading from WWTPs over the period 1950–2015. The ensemble reflect the uncertainty in the characteristics of different parameters, such as the fraction of protein supply lost in the food supply chain, the ratio of industrial to domestic N emissions or the efficiency of wastewater treatment. Supplementary Section S6 (as well as Figure S4 and Tables S1-S2) details the underlying procedure for the N point sources construction. A visual depiction of the N point sources with uncertainty is provided in Figure S5.

3.2.5 Stream Discharge

To run the ELEMeNT model, we require annual discharge at the outlet of the sub-catchments for the period 1800–2015 and, in addition, we need daily discharge for the recent period to process the measurements of in-stream N-NO_3 concentration, as described in Section 3.2.6. Discharge data is constructed by combining (1) daily discharge measurements (at the catchment outlet or at a nearby measuring station) obtained from the GRDC (Global Runoff Data Centre, 2021) or the WRB Commission (FGG Weser, 2021) databases (see supplementary Table S3), and (2) bias-corrected simulations from the mesoscale Hydrologic Model (mHM, Kumar et al., 2013; Samaniego et al., 2010), to fill the missing values in the observation dataset. Two sets of mHM simulations are used, medium-term daily simulations for the period 1950–2015 (Zink et al., 2017), and long-term annual simulated values for the period 1800–1949 (Hanel et al., 2018). The mHM simulations capture the variability of observed discharge reasonably well, with values of the Nash–Sutcliffe efficiency always higher than 0.64 and values of the coefficient of determination always higher than 0.66 (see supplementary Table S4). Figures S6–S13 represent the annual time series of the discharge measurements and the simulations before and after we apply the bias-correction. We refer to Zink et al. (2017) and Hanel et al. (2018) for details on the mHM setups.

3.2.6 Observations of N Content

N content in the Source Zone

We derive N content in the topsoil (0–20 cm) from the Land Use and Cover Area frame statistical Survey (LUCAS; Ballabio et al., 2016, 2019). The LUCAS dataset was created from the spatial interpolation of approximately 22,000 surveyed points across Europe. For most countries, including Germany, soil samples were collected in 2009. To estimate the catchment-scale soil N content (0–100 cm, source zone of ELEMeNT), we combine the N content and bulk density of the topsoil from the LUCAS dataset. We also use the ratio of total soil N content (0–100 cm) to topsoil N content (0-20 cm), which we estimate to be between 2.5 and 4 from Batjes (1996), thus obtaining a plausible range for the total soil N content (0–100 cm). Our estimated ranges of the soil N content and further details are reported in supplementary Table S5.

In-stream N Concentration

In-stream nitrate concentration is obtained from the WRB Commission (FGG Weser, 2021). For the Letzter Heller catchment, we combine the concentration measurements at the Letzter Heller station available for the period 1979–2002 and at the Witzenhausen station, which is located 8 km upstream, for the period 2003–2015. The data consists of 14-day average N-NO_3 concentration measurements, that were constructed by mixing daily samples, and that start for most stations in the early 1980s and end around 2015. We only fill the gaps that have a length of maximum 42 days, and therefore concentration data for years containing longer gaps are discarded. Annual total in-stream loading is calculated as the sum of the 14-day average loading values, while annual average concentration is estimated as the discharge weighted average of the 14-day average concentration values. The in-stream N-NO_3 concentration and loading data are reported in supplementary Figures S14–S15. We also examine the uncertainty in the observations and check for outlier values. We find that the concentration value at the Letzter Heller station in 1990, which is equal to 6.3 $mg\ L^{-1}$, is abnormally high. The difference to the average value (4.5 $mg\ L^{-1}$) amounts to 2.8 times the standard deviation over the period 1985–1995 (see supplementary Figure S14). This anomalous concentration value could be explained in the context of the German reunification in 1990, where unusual and undocumented N amounts could have been discharged into the stream. This is not reflected in our N input datasets. Tracking the cause of this anomaly is beyond the scope of this study, and therefore, we discard this value for our analyses.

4 Methods for Parameter Estimation and Sensitivity Analysis

4.1 Multicriteria Parameter Estimation Strategy

Our parameter estimation strategy considers the performance of the ELEMeNT model in simulating three different variables, namely the total source zone N content (M_s, which includes the organic protected, organic active, and inorganic N stores), and the in-stream N-NO_3 loading (J_{out}) and concentration (C_{out}) at the catchment outlet. Regarding M_s, the simulated values are constrained within the range derived from the observations, since the source zone N content is only provided for the year 2009. For J_{out} and C_{out}, we use three performance metrics that are the Pearson correlation coefficient denoted as ρ (-), the relative bias denoted as $RBIAS$ (-) and the variability error denoted as STD_{err} (-). $RBIAS$ and STD_{err} are defined as follows:

$$RBIAS = \frac{\mu_{sim} - \mu_{obs}}{\mu_{sim}} \tag{5}$$

$$STD_{err} = \frac{\sigma_{sim} - \sigma_{obs}}{\sigma_{sim}} \tag{6}$$

where μ_{obs} and μ_{sim} are the average of the observations and simulations, respectively, and σ_{sim} and σ_{obs} are the standard deviation of the simulations and observations, respectively. There averages and standard deviations are calculated for each subcatchment over the years where observations are available. The three metrics ρ, $RBIAS$ and STD_{err} measure how well the dynamics (temporal pattern and timing), the mean and the variability of the observations respectively are captured by the simulations. They constitute the three components of the Kling-Gupta efficiency (*KGE;* Gupta et al., 2009) defined as:

$$KGE = 1 - \sqrt{(1-\rho)^2 + RBIAS^2 + STD_{err}^2} \tag{7}$$

We consider these three metrics separately instead of the aggregated KGE measure to ensure a sufficient performance with regards to all important aspects that we aim to simulate, as discussed in Martinez and Gupta (2010).

Similar to e.g., Choi and Beven (2007), Hartmann et al. (2017), (Husic et al., 2019), and Sarrazin et al. (2018), we use "soft rules" to identify the set of well-performing ("behavioural") simulations. We define seven soft rules, all of which have to be satisfied in the behavioural simulation ensemble:

1. for J_{out}: $\mid RBIAS \mid \leq 0.2$;
2. for J_{out}: $\mid STD_{err} \mid \leq 0.25$;
3. for J_{out}: $\rho \geq 0.8$;
4. for C_{out}: $\mid RBIAS \mid \leq 0.2$;
5. for C_{out}: $\mid STD_{err} \mid \leq 0.25$;
6. for C_{out}: $\rho \geq 0.6$;
7. the simulated M_s is within the range derived from the observations.

The order of the rules allows us to assess to what extent the use of in-stream N concentration and source zone N content data can help to reduce the simulation uncertainties, beyond the use of in-stream N loading data. Some studies only examine the in-stream N loading (e.g. Chang et al., 2021; J. Liu et al., 2021; Van Meter et al., 2017, 2018) and not the in-stream N concentration that tends to be more difficult to simulate than the in-stream N loading (Husic et al., 2019). In addition, as discussed in Section 1, previous studies generally considered in-stream variables only for calibration. The threshold values

for $RBIAS$ and ρ introduced in rules 1, 3, 4 and 6 correspond to "satisfactory" or "good" model performance in reproducing nutrient dynamics according to Moriasi, Gitau, et al. (2015). We however note that Moriasi, Gitau, et al. (2015) examine values of ρ at monthly and not annual timescale, due to a lack of studies analysing the model goodness-of-fit for annual simulations of nutrients. Here, we set a stricter threshold value on ρ for J_{out} (rule 3) compared to C_{out} (rule 6), since the dynamics of J_{out} are driven by the stream discharge and are therefore easier to reproduce then the dynamics of C_{out}, as further discussed in Sections 5.1.1-5.1.2. Due to a lack of analysis of STD_{err} in previous studies, we consider that a threshold value equal to $\pm$ 0.25 is reasonable (rules 2 and 5).

To estimate the nine model parameters, we generate a parameter sample of size 100,000 from the ranges reported in Table 1 utilizing latin hypercube sampling and uniform distributions. We discard the parameter sets that do not meet the condition $h_c < h_{nc}$ (where h_c and h_{nc} are the protection coefficients for cultivated and non-cultivated land, respectively). We assume that the protection of organic matter is reduced by tillage practices. We perform Monte-Carlo simulations for each of the eight subcatchments and we sequentially apply the seven soft rules, thus progressively reducing the number of behavioural parameter sets. Following previous studies (Dupas et al., 2020; Ehrhardt et al., 2021; Van Meter et al., 2017), we use the entire time series of in-stream N observations to identify behavioural simulations. We note that the goal of our analysis here is not to predict future catchment N export but to analyze the uncertainty in the simulations and the value of different types of data to constrain the simulations. In addition, the amount of in-stream observational data available (between 26 and 37 years depending on the subcatchments) is rather limited to be divided between a calibration set and an independent verification set.

We simulate the ELEMeNT model from year 1800 (pre-industrial conditions) to year 2015, including a long warm-up period i.e., we only analyze the simulations for the period 1960–2015 and we discard the results for period 1800–1959. This is because the legacy stores can have a slow turnover and can build up over long timescales (Van Meter et al., 2017). The setup of the initial states is described in detail in supplementary Section S7. Regarding the N point sources, we select the realization in our ensemble that shows the best match with the observations available for the period 2012–2016 (further information on the point sources are in Section 3.2.4). For the atmospheric N deposition, we use the value in 1850 for the period 1800–1849. Regarding the N surplus in agricultural areas, since no data are available before 1950, we assume that the value (at county level) of the N surplus in 1850

is half the value in 1950. We then use linear interpolation for the period 1850–1950 and we consider that the value is constant for the period 1800–1850. We also assume that the N surplus takes the same value over cropland and agricultural permanent grassland. In Section 4.2, we explain how we assess the impact of the uncertainty in the N point sources and N surplus.

4.2 Sensitivity Analysis of the Simulated N Legacies

We perform a sensitivity analysis to investigate the factors that are responsible for the residual uncertainty in the simulated N legacy stores, i.e., the uncertainty that remains after constraining the model simulations using the soft rules described in Section 4.1. This analysis allows us to set priorities for future efforts for uncertainty reduction and model improvement. We examine the sensitivity of four model outputs related to the legacy stores (namely the average source zone and subsurface storage and their cumulative change, assessed over the period 1960–2015) to the nine parameters of ELEMeNT, the N point sources input and the N surplus input. We select ten realizations of the N point sources across the ensemble of realizations to cover the uncertainty range (see Figure S5). Regarding the N surplus, we introduce two additional parameters to generate alternative realizations for the agricultural N surplus for the period (1800–1949) and for the disaggregation between cropland and agricultural permanent grassland. First, we define parameter r_{warm} (-), which represents the ratio of the value in 1850 to the value in 1950 of the agricultural N surplus $Surplus_{agr}(t)$ $(kg\ ha^{-1}\ yr^{-1})$:

$$r_{warm} = \frac{Surplus_{agr}(1850)}{Surplus_{agr}(1950)} \tag{8}$$

The agricultural N surplus for the period 1800–1949 is derived from r_{warm}, through linear interpolation over the period 1850–1950 and by setting a constant value over the period 1800–1850. Second, we define parameter $r_{mgra-crop}$ (-), which is the ratio of the N surplus for permanent agricultural grassland $Surplus_{mgra}(t)$ $(kg\ ha^{-1}\ yr^{-1})$ to the N surplus for cropland $Surplus_{crop}(t)$ $(kg\ ha^{-1}\ yr^{-1})$, assumed to be constant in time:

$$r_{mgra-crop} = \frac{Surplus_{mgra}(t)}{Surplus_{crop}(t)} \tag{9}$$

In addition, we define a time-invariant multiplier denoted as $f_{surplus}$ (-), which is used to multiply the time series 1800–2015 of the N surplus for both agricultural and non-agricultural areas. This multiplier accounts for the uncertainty in the value of the total N surplus.

We select three values for $f_{surplus}$ (0.8, 1 and 1.2), r_{warm} (0.25, 0.5 and 0.75) and $r_{mgra-crop}$ (0.5, 1 and 1.5), which results in 27 N surplus realizations. Since no information is available to further constrain the uncertainty, we argue that these 27 realizations cover a large plausible range of N-surplus estimates. The case $f_{surplus} = 1$, $r_{warm} = 0.5$ and $r_{mgra-crop} = 1$ corresponds to our "baseline scenario" i.e., the one used for analyses presented in Section 4.1. Supplementary Section S8 details the derivation of the N surplus for cropland and permanent agricultural grassland from the parameter $r_{mgra-crop}$ and supplementary Figure S16 reports the time series of different realizations of the N surplus for agricultural areas.

We combine the ten point sources and the 27 N surplus realizations to create 270 sets of N inputs. For each of them, we perform Monte-Carlo simulations from the same parameter sample of size 100,000 described in Section 4.1. This produces a total input-output sample of size 27,000,000. We then discard simulations that do not satisfy the soft rules defined in Section 4.1 to obtain the sample for the sensitivity analysis.

We apply the distribution-based PAWN sensitivity analysis method (Pianosi & Wagener, 2015) that evaluates the effect of the input factors on the entire output (here legacy stores) distribution. We estimate the PAWN sensitivity indices using the numerical approximation strategy introduced by Pianosi and Wagener (2018), which can be utilized for any generic input-output sample, and which is implemented in the Python version of the SAFE toolbox (Pianosi et al., 2015). With this numerical scheme, the range of variation of the i-th input factor is partitioned into a number n_i of equally probable "conditioning" intervals (denotes as $I_{i,k}$, $k=1,...,n_i$), each interval containing the same number of parameter sets. The PAWN method consists of the comparison between 1) the Cumulative Distribution Functions (CDFs) of the model output (denoted here as y) obtained by letting all input factors vary in their entire space of variability (i.e., unconditional CDF, denoted as $F_y(y)$) and 2) the CDF obtained by allowing all input factors to vary freely, but the i-th input x_i whose value is constrained to a specific conditioning interval (i.e., conditional CDF, denoted as $F_{y|x_i}(y|x_i)$). In PAWN, input sensitivity is quantified through the Kolmogorov-Smirnov statistic (KS, Kolmogorov, 1933; Smirnov, 1939), which is the maximum vertical distance, between unconditional and conditional CDFs. The PAWN sensitivity index for the i-th input factor, denoted as S^i_{PAWN} (-), aggregates the KS values calculated across all n_i conditioning intervals through a summary statistic, which is chosen as the median value in this

study, to eliminate the impact of outlier values:

$$S^i_{PAWN} = \underset{x_i}{Q_{50}}\ KS(x_i) \tag{10}$$

where

$$KS(x_i) = \max_y |F_y(y) - F_{y|x_i}(y|x_i)| \tag{11}$$

S^i_{PAWN} takes values between 0 and 1, and the higher its value the larger the impact of that input on the model output. For the ELEMeNT parameters, we adopt a number of conditioning intervals n_i equal to 10. For the N point sources, we calculate the conditional CDF for each of the ten realizations, and for the three N surplus parameters we compute the conditional CDF for each of their three selected values. We estimate the 95% confidence intervals of the PAWN sensitivity indices using 1000 bootstrap resamples, and we verify the convergence of the results given the sample size, following Sarrazin et al. (2016).

5 Results

5.1 Parameter Estimation

5.1.1 Application of the Soft Rules

The performance of the simulated in-stream N loading and concentration is comparable in terms of the metrics $RBIAS$ and STD_{err} (Figure 2a). However, in terms of ρ, the performance is noticeably better for the simulated N loading compared to the concentration for five subcatchments. For the three other subcatchments (outlet Hemelingen, Drakenburg and Wahnhausen), performance are more similar although higher values of ρ can be reached for the loading. In contrast to the concentration dynamics, the temporal fluctuations of the loading are strongly influenced by the discharge dynamics, which is an input to the ELEMeNT model. Importantly, we identify simulations that comply with each soft rule individually, as shown in the grey shaded areas in Figure 2a for rules 1-6 and in supplementary Figure S17 for rule 7 (source zone N content). We also find that the threshold values on the performance metrics introduced in rules 1-6 result in values of the KGE higher than 0.62 for the loading and 0.49 for the concentration in the behavioural simulation ensemble (grey shaded areas in right panels of Figure 2a). We verify that these KGE values are higher than the mean benchmark value of -0.41 (Knoben et al., 2019).

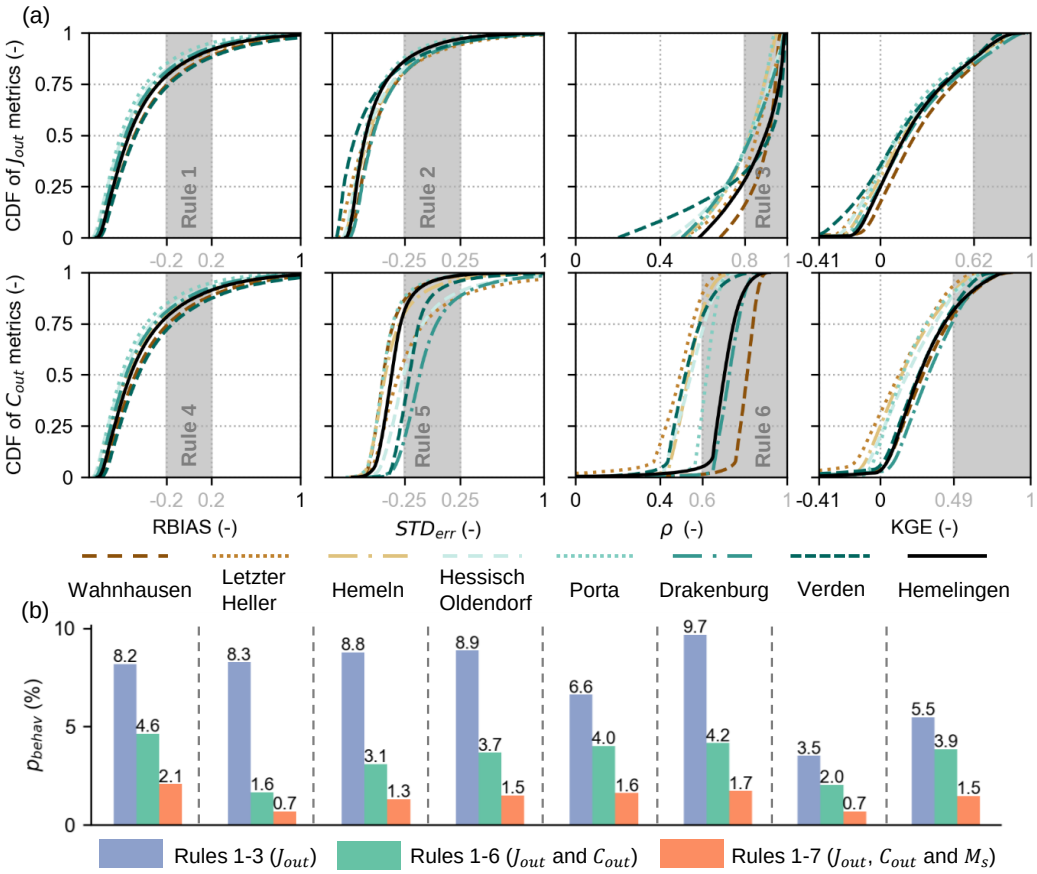

Figure 2. Application of the soft rules: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream N loading (J_{out}) and concentration (C_{out}) in the initial simulation ensemble (100,000 realizations) and (b) percentage of realizations of the initial ensemble identified as behavioural (p_{behav}) by successive application of the soft rules based on the performance metrics for loading (J_{out}, rules 1-3), the performance metrics for concentration (C_{out}, rules 4-6), and the source zone N content (M_s, rule 7). The name of the eight subcatchments refer to both the legend of the lines of panel (a) and the bar graphs of panel (b). Panel (a) reports the three performance metrics used in the definition of rules 1-6 (relative bias $RBIAS$, variability error STD_{err} and Pearson correlation coefficient ρ) and the Kling-Gupta efficiency (KGE). The grey shaded areas and grey numbers on the x-axis indicate the behavioural ranges of the performance metrics used in the definition of rules 1-6. The ranges of the performance metrics shown (x-axis of plots in panel a) do not include the extreme values, which are shown in supplementary Figure S18.

From Figure 2b, we observe a reduction in the number of ELEMeNT realizations after application of the rules on the loading (rules 1-3), but also a further diminution after applications of the rules on the concentration (rules 4-6) and on the source zone N content (rule 7). This means that not only the loading, but also the concentration and the source zone N content have a value in constraining the simulations. In addition, the data on the source zone N content in 2009 (rule 7) allows reduction in the uncertainty in the total source zone N storage that is not constrained by the other rules (Figure S17). We obtain a number of behavioural simulations that varies between 676 for Letzter Heller and 2076 for Wahnhausen (supplementary Table S6). For further details on the reduction in the number of realizations obtained when applying each of the seven rules individually, we refer to supplementary Figure S19.

5.1.2 Constraining of the Simulated In-stream Loading and Concentration

From Figures 3-4, we observe that the precision of the simulated in-stream N loading and concentration is larger in the behavioural simulation ensemble compared to the unconstrained ensemble, i.e., the red shaded areas are much narrower than the grey shaded areas. We also see that, for some years, the width of the 95% confidence interval (CI) of the simulation ensemble is reduced when applying the rules on the concentration and source zone N content (rules 4-7, red shaded areas) in addition to the rules on the loading (rules 1-3, blue shaded areas). However, the information on the source zone N content in 2009 (rule 7) does not further narrow down the uncertainty bounds, i.e., we do not observe green shaded areas. Importantly, after applying the seven rules, the behavioural simulation ensembles (red shaded areas in Figures 3-4) capture a large number of observations. Specifically for most subcatchments, the simulation ensemble encompasses more than 90% of the observations, except for Verden for which it includes 57% of the observations.

Regarding loading (Figure 3), the temporal dynamics follow the discharge dynamics (discharge is reported in supplementary Figures S6–S13), and we see that the simulation ensembles match the observations very well. Regarding concentration (Figure 4), we observe that the simulations are overall in agreement with the measurements. However, the median ensemble has difficulties in reproducing the observed concentration trend around the last ten years of the simulations (2005–2015), in particular at Hemeln, Porta, Verden and Hemelingen. While the measurements indicate a slight decrease in concentration, the latter is relatively stable in the simulations, which is consistent with the N input dynamics

over this time period (see Figure 1). Nonetheless, the simulation ensemble captures the observations for all locations apart from Verden. We also notice that for Letzter Heller (panel b), concentration shows little temporal variations, while the N surplus trajectory presents a sharp decrease in 1990 (as shown in Figure 1).

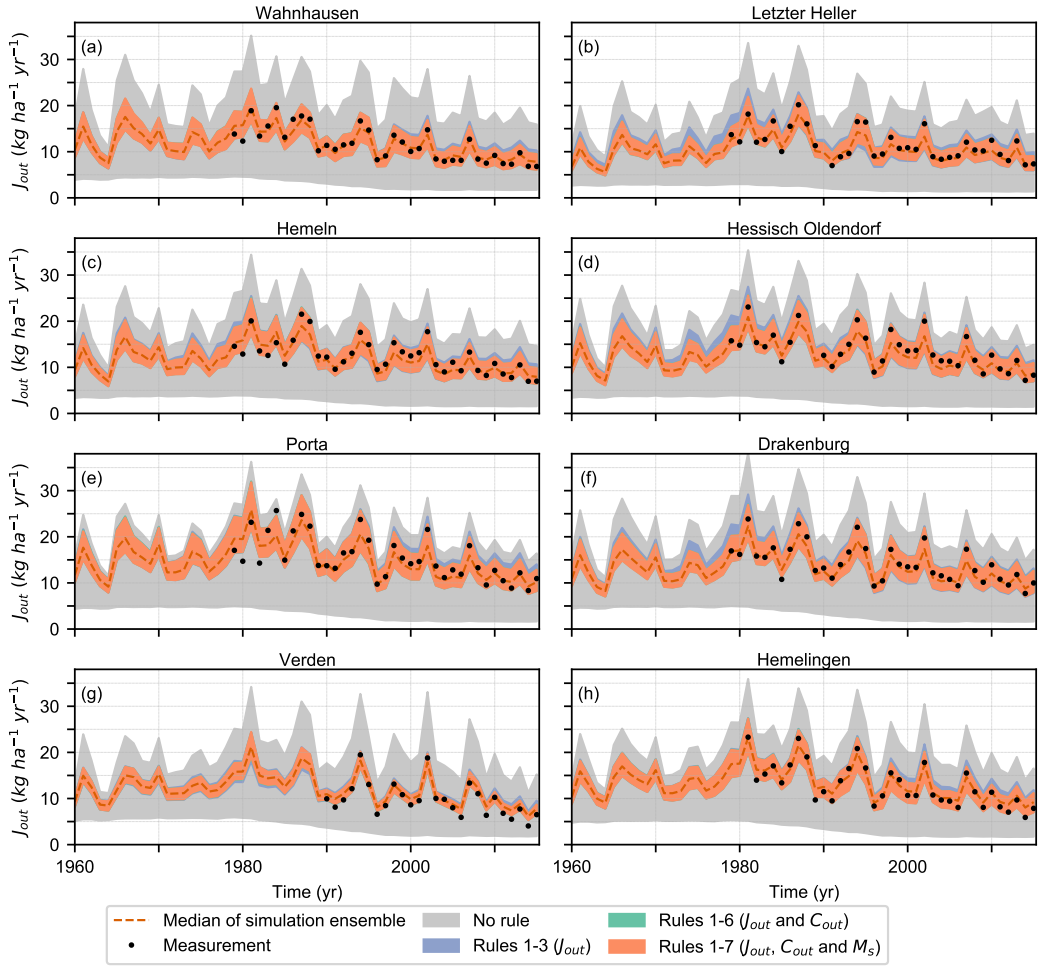

Figure 3. Simulated annual in-stream N-NO_3 loading (J_{out}) for the eight subcatchments. The dashed red lines represent the median of the behavioural simulation ensembles that satisfy all seven soft rules. The shaded areas represent the 95% CI of the simulation ensembles corresponding to four different levels of constraining.

5.1.3 Constraining of the Parameter Distributions

From Figure 5, we find that the differences between the prior (grey lines) and posterior (colored lines) CDFs of the parameters are very small for four parameters, namely the two

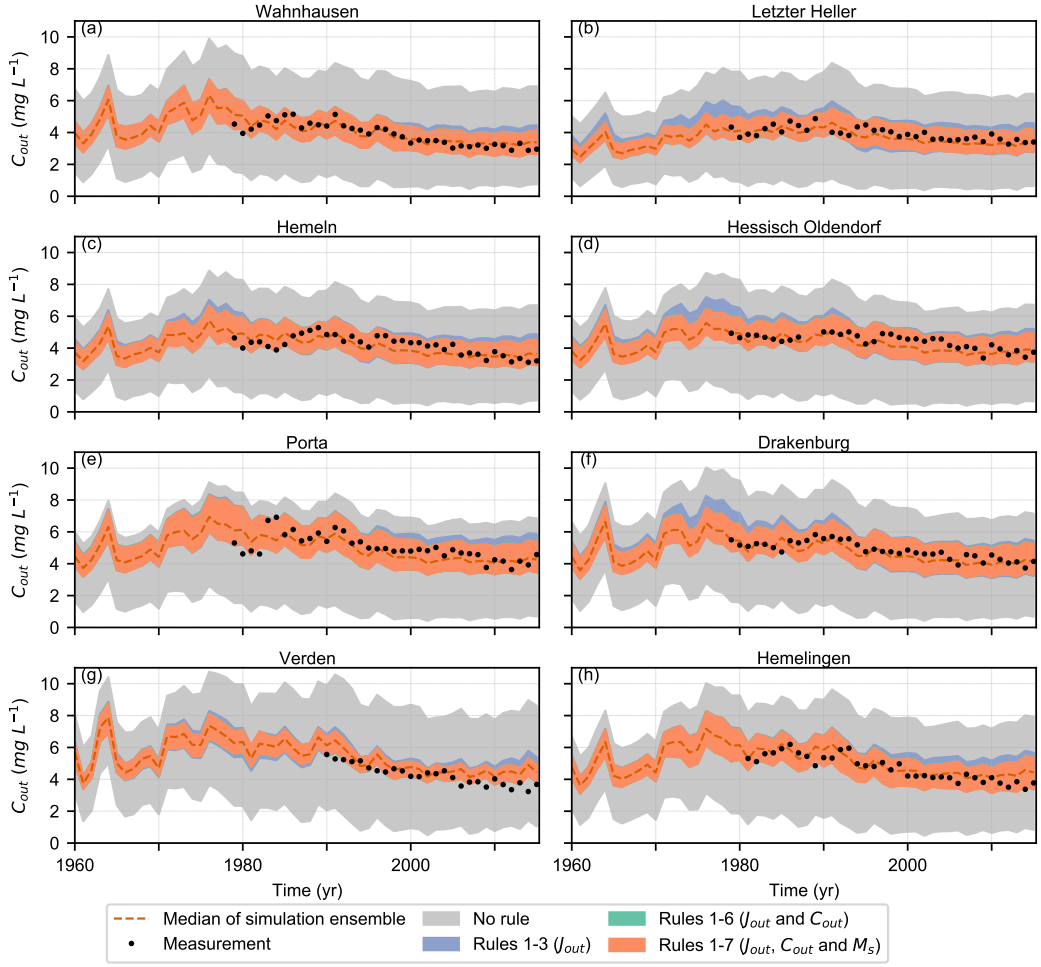

Figure 4. Simulated annual in-stream N-NO_3 concentration (C_{out}) for the eight subcatchments. The dashed red lines represent the median of the behavioural simulation ensembles that satisfy all seven soft rules. The shaded areas represent the 95% CI of the simulation ensembles corresponding to four different levels of constraining.

protection coefficients (h_c and h_{nc}), the mean annual water content in the source zone (V_s) and the fraction of in-stream N removal (R). Nonetheless, the median of the distribution changes by at least 20% for V_s for one subcatchment (increase of 25% for Letzter Heller) and for R for four subcatchments (decrease of 27% for Wahnhausen, 20% for Hemeln, 22% for Porta and 23% for Hemelingen). The other five parameters ($M_{s_{org}}^{prist}$, k_a, λ_s, λ_{sub}, and μ_{sub}) show appreciable (higher than 15%) reduction in their 95% CI. On average, these parameters exhibit a diminution in their CI of 59%, 10%, 37% 15% and 33%, respectively. The two denitrification rate constants (λ_s and λ_{sub}) and the mean travel time (μ_{sub}) take

values in the lower range of their prior distribution for most subcatchments, with median values in the range 0.28–0.34 yr^{-1} (prior: 0.55 yr^{-1}), 0.06–0.17 yr^{-1} (prior: 0.15 yr^{-1}) and 8–19 yr (prior: 26 yr), respectively. The 95% CI of the travel time in the posterior distribution is reduced for the some subcatchments such as Hemelingen (95% CI: 2–24 yr), while it can still be rather large for some other subcatchments such as Drakenburg (95% CI: 3–44 yr). The mineralization rate constant for organic active N (k_a) is unlikely to take values in the lower range of its prior distribution, as the lower bound of its 95% CI is in the range 0.09–0.18 yr^{-1} (prior: 0.07 yr^{-1}). Details on the median and 95% CI of the parameter values are reported in supplementary Tables S7–S8. In addition, all three observational data used (loading, concentration and source zone N content) have value in constraining the distribution of at least one parameter, as shown in supplementary Figures S20–S27. In particular, the source zone N content observations in 2009 is the only data source that allows to constrain the source zone organic N stock under pristine conditions ($M_{s_{org}}^{prist}$).

We also calculate the behavioural values of the mean transfer times for the source zone organic N stores as the inverse of the respective mineralization rate constants. The mineralization rate constant for the protected pool is computed from Equation S24 and its CDFs can be visualized in supplementary Figure S28. We find that the mean transfer times for the protected store are much higher compared to the active store for all subcatchments. The median values (95% CI) of the mean transfer times are in the range 2000–2700 yr (1300–5400 yr) and 2.0–3.0 yr (1.4–11.3 yr) for the protected and active stores, respectively (more details in supplementary Table S9).

Interestingly, the parameter distributions for the upstream Letzter Heller subcatchment stands out, with particularly low median values of k_a, and high median values of μ_{sub} compared to the other subcatchments, including the Wahnhausen subcatchment which is neighbouring the Letzter Heller and which has similar land use and topography (Figure 1, Section 3.1). The results suggest that, compared to the other subcatchments, Letzter Heller may have a particularly high potential to accumulate organic active N in the source zone, due to the low mineralization rate for the active N pool (median value: 0.33 yr^{-1}), and dissolved mineral N in the subsurface, due to the large travel time (median value: 17 yr).

Overall, our results demonstrate that, for each of the eight subcatchments, multiple combinations of model parameter values lead to acceptable model performance. This equifinality can be partly explained by interactions between the model parameters, in particular

between the two denitrification rate constants in source zone and subsurface, between each of the denitrification rate constant and the mean travel time in the subsurface, between the denitrification rate constant in the subsurface and the fraction of in-stream removal and between the two protection coefficients (a detailed interaction analysis is presented in Table S10).

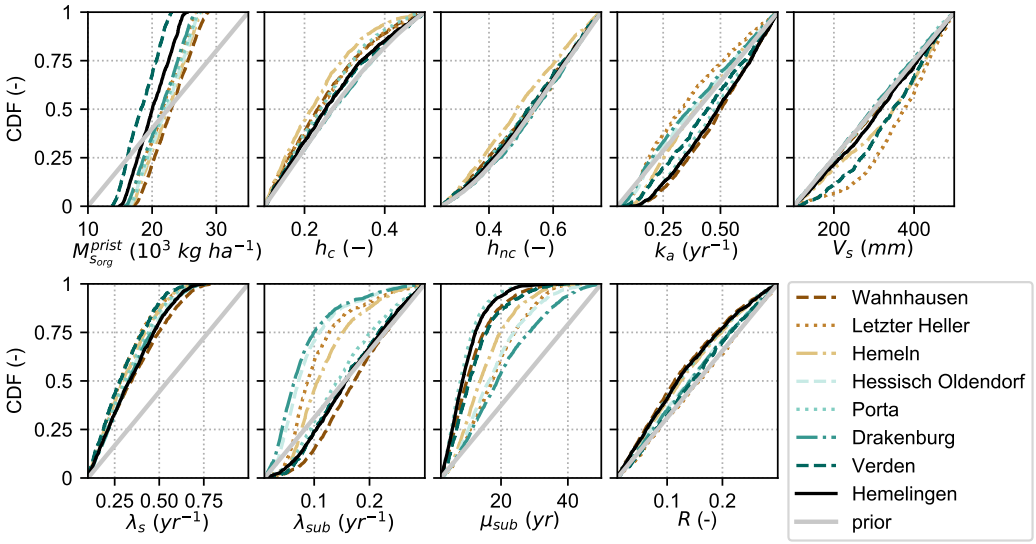

Figure 5. Cumulative distribution functions (CDFs) of the model parameters for the eight subcatchments. Colored lines refer to the behavioural CDFs, which are obtained after applications of the seven soft rules, and grey lines refer to the prior CDFs in the original sample of size 100,000. The prior CDFs are the same for all subcatchments.

5.2 Mass balance Over the Study Domain

5.2.1 Fate of the N Surplus

In this section, we examine the fate of the N surplus over the period 1960–2015 from the behavioural simulation ensemble (Figure 6, Table 2). Over the entire WRB (Hemelingen), the total denitrification in the source zone and subsurface ($J_{den_{tot}}$) amounts to 1888 $kg\ ha^{-1}$ (median value), which corresponds to about 63% of the N surplus (95% CI: 49–74%). Around 50% (median value) of the total denitrification occurs in the source zone, but the uncertainty in the partitioning of denitrification between source zone and subsurface is large (95% CI of the source zone contribution: 18–94%; supplementary Table S11).

667 The landscape export of N from catchment ($J_{out_{sub}}$) is equal to 537 $kg\ ha^{-1}$ (median
668 value), which represents 18% of the N surplus. As the initial source zone N storage in 1960
669 is very large (median value: 17,487 $kg\ ha^{-1}$; 95% CI: 13,678–21,564 $kg\ ha^{-1}$), the change
670 in the source zone N storage (ΔM_s, biogeochemical legacy) is relatively small with respect
671 to this initial storage (median value: 2.5%; 95% CI: 1.2-5.1%; see details in supplementary
672 Figure S30 and Table S12). Nevertheless, the change in the source zone storage amounts to
673 448 $kg\ ha^{-1}$, (median value), which constitutes a high percentage of the N surplus (15%)
674 similar to the stream export part. The 95% CI is however large for this component (229–781
675 $kg\ ha^{-1}$) and is largely overlapping with the 95% CI of the stream export. The two last
676 components per order of magnitude are the in-stream removal ($J_{rem_{sub}}$) and the change in
677 the subsurface N storage (ΔM_{sub}, hydrologic legacy), which have median values equal to 72
678 and 37 $kg\ ha^{-1}$ respectively (which correspond to 2% and 1% of the N surplus respectively),
679 and which have overlapping 95% CI. Therefore, the denitrification in the source zone and
680 subsurface is an order of magnitude greater than the in-stream removal. Moreover, the N
681 accumulation in the source zone is an order of magnitude higher than the N accumulation
682 in the subsurface. Total legacy buildup in the WRB amounts to 491 $kg\ ha^{-1}$ (95% CI:
683 264–820 $kg\ ha^{-1}$), which corresponds to around 16% of the N surplus.

684 We also observe that the simulated source zone N store is continuously building up
685 in time over the period 1960–2015 (Figure 6a). In the subsurface, the dynamics of the N
686 store is much more coupled to the dynamics of the N surplus (Figure 6b). We see that
687 the N storage in the subsurface increases until the 1987 to reach a value of 61 $kg\ ha^{-1}$
688 (median value), decreases by about as much as 40% between 1987 and 2010, and shows
689 small fluctuations between 2010 and 2015.

690 For the different subcatchments, the relative importance of the different components of
691 the N mass balance are similar (Table 2). In particular, denitrification $J_{den_{tot}}$ is always the
692 largest outgoing N flux. The median change in source zone storage ΔM_s generally varies
693 between 473 $kg\ ha^{-1}$ (Hemeln) and 584 $kg\ ha^{-1}$ (Drakenburg). An exception is the Verden
694 subcatchment, which is mostly located in the lowland areas, and for which the median ΔM_s
695 is smaller (326 $kg\ ha^{-1}$). The median change in subsurface storage ΔM_{sub} is smaller for
696 Wahnhausen (17 $kg\ ha^{-1}$) and is larger for Letzter Heller (62 $kg\ ha^{-1}$), Hessisch Olden-
697 dorf (62 $kg\ ha^{-1}$) and Drakenburg (78 $kg\ ha^{-1}$). The relatively large value of ΔM_{sub} for
698 Letzter Heller compared to the other subcatchments is consistent with the parameter dis-
699 tribution results presented in Section 5.1.3. We also observe that the (temporal) dynamics

of N buildup in the legacy stores of all subcatchments are similar to those of Hemelingen (see supplementary Figures S29-S31). In particular, N levels in the subsurface peak around the year 1990. Notably, the highest level of N accumulation in the subsurface across subcatchments and time is equal to 119 $kg\ ha^{-1}$ (median value) and is reached for Drakenburg in 1993 and Letzter Heller in 1987. However, the differences found between catchments are not robust, since the 95% CI are largely overlapping between subcatchments.

We also examine the change in the different N stores of the source zone, i.e., the organic protected, organic active, and inorganic N stores (details in supplementary Table S12). Most of the N accumulation occurs in the protected pool (e.g. 94% for Hemelingen; 95% CI: 79–98%). For Hemelingen, N buildup amounts to 448 $kg\ ha^{-1}$ (95% CI: 229–781 $kg\ ha^{-1}$) in the protected store, to 21 $kg\ ha^{-1}$ (95% CI: 12–74 $kg\ ha^{-1}$) in the active store, and to 2 $kg\ ha^{-1}$ (95% CI: -5–16 $kg\ ha^{-1}$) in the inorganic store.

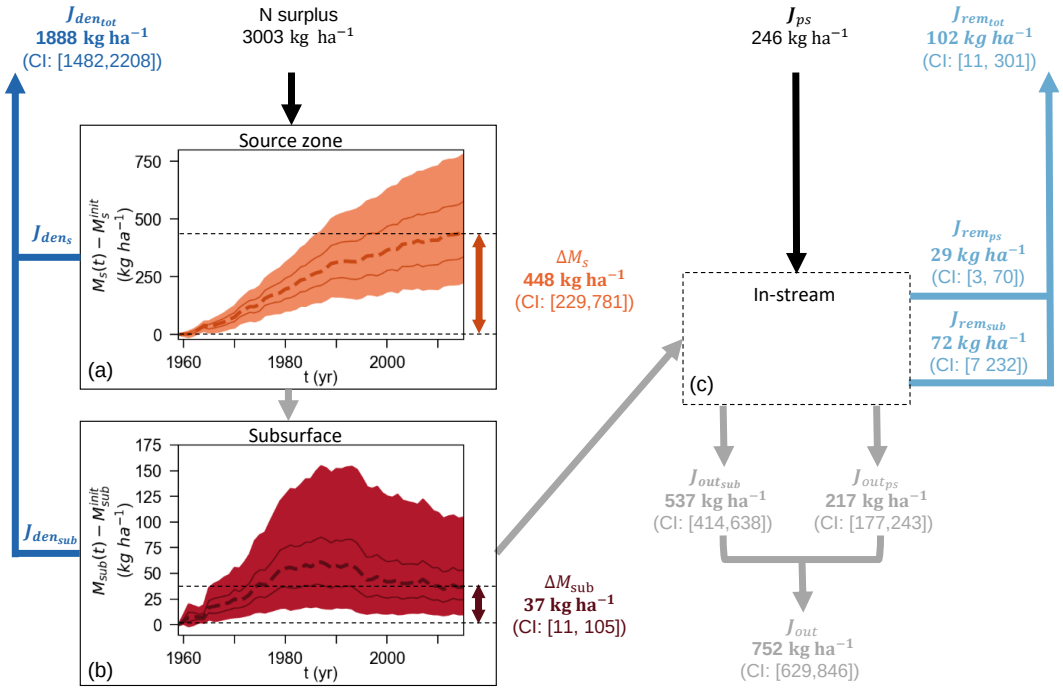

Figure 6. Cumulative values of the components of the N mass balance (inputs and simulated variables) for the WRB at Hemelingen for the period 1960–2015. For the simulated variables, the figure reports the median values and the 95% CI in the behavioural simulation ensemble. Panels (a-b) report the simulated cumulative change in N storage for the source zone (M_s) and the subsurface zone (M_{sub}) since 1960 as a function of time t (M_s^{init} and M_{sub}^{init} are the initial conditions in 1960 for the source zone and subsurface storage respectively). The shaded areas indicate the 95% CI, the solid lines the 25% and 75% quantiles and the dashed lines the median values. Panel (c) represents the in-stream compartment, where no accumulation of N occurs in ELEMeNT. Notations: J_{ps} is the point source; J_{out} is the simulated in-stream loading, which is the sum of the point source contribution ($J_{out_{ps}}$) and the subsurface contribution ($J_{out_{sub}}$); $J_{den_{tot}}$ is the total denitrification, which is the sum of the denitrification in the source zone (J_{den_s}) and in the subsurface ($J_{den_{sub}}$); $J_{rem_{tot}}$ is the in-stream removal, which is the sum of the removal of the point source contribution ($J_{rem_{ps}}$) and the subsurface contribution ($J_{rem_{sub}}$); ΔM_s is the change in the source zone storage which includes three N stores (organic protected, organic active, and inorganic N stores); ΔM_{sub} is the change in the subsurface storage.

Table 2. Components of the N Mass Balance in the Behavioural Simulation Ensemble for the Period 1960–2015

Variable	Wahn-hausen	Letzter Heller	Hemeln	Hessisch Olden-dorf	Porta	Draken-burg	Verden	Heme-lingen
	($kg\ ha^{-1}$)							
Surplus	3093	2917	3004	3041	3075	3179	2697	3003
$J_{out_{sub}}$	477^{604}_{361}	461^{586}_{382}	513^{645}_{404}	550^{704}_{455}	632^{813}_{510}	553^{706}_{444}	446^{500}_{392}	537^{638}_{414}
$J_{den_{tot}}$	2001^{2315}_{1616}	1805^{2096}_{1473}	1872^{2169}_{1501}	1787^{2124}_{1412}	1787^{2128}_{1403}	1838^{2214}_{1436}	1786^{2088}_{1436}	1888^{2208}_{1482}
$J_{rem_{sub}}$	62^{214}_{7}	88^{212}_{8}	72^{224}_{7}	91^{249}_{9}	87^{276}_{8}	91^{250}_{9}	75^{192}_{8}	72^{232}_{7}
ΔM_s	504^{828}_{290}	480^{764}_{270}	473^{786}_{283}	526^{803}_{278}	485^{807}_{283}	584^{911}_{306}	326^{642}_{80}	448^{781}_{229}
ΔM_{sub}	17^{54}_{5}	62^{138}_{15}	39^{100}_{10}	62^{160}_{16}	30^{90}_{9}	78^{206}_{17}	44^{142}_{11}	37^{105}_{11}
J_{ps}	214	144	183	180	226	226	274	246
$J_{out_{ps}}$	190^{211}_{153}	121^{141}_{102}	161^{181}_{131}	154^{176}_{127}	199^{223}_{162}	194^{222}_{161}	234^{269}_{194}	217^{243}_{176}
$J_{rem_{ps}}$	24^{61}_{3}	23^{42}_{3}	23^{52}_{3}	25^{52}_{3}	27^{64}_{3}	32^{66}_{4}	40^{80}_{5}	29^{70}_{3}
	(% N input)							
Surplus	100	100	100	100	100	100	100	100
$J_{out_{sub}}$	15^{20}_{12}	16^{20}_{13}	17^{21}_{13}	18^{23}_{15}	21^{26}_{17}	17^{22}_{14}	17^{19}_{15}	18^{21}_{14}
$J_{den_{tot}}$	65^{75}_{52}	62^{72}_{51}	62^{72}_{50}	59^{70}_{46}	58^{69}_{46}	58^{70}_{45}	66^{77}_{53}	63^{74}_{49}
$J_{rem_{sub}}$	2^{7}_{0}	3^{7}_{0}	2^{7}_{0}	3^{8}_{0}	3^{9}_{0}	3^{8}_{0}	3^{7}_{0}	2^{8}_{0}
ΔM_s	16^{27}_{9}	16^{26}_{9}	16^{26}_{9}	17^{26}_{9}	16^{26}_{9}	18^{29}_{10}	12^{24}_{3}	15^{26}_{8}
ΔM_{sub}	1^{2}_{0}	2^{5}_{1}	1^{3}_{0}	2^{5}_{1}	1^{3}_{0}	2^{6}_{1}	2^{5}_{0}	1^{3}_{0}
J_{ps}	100	100	100	100	100	100	100	100
$J_{out_{ps}}$	89^{99}_{71}	84^{98}_{71}	88^{99}_{71}	86^{98}_{71}	88^{99}_{72}	86^{98}_{71}	85^{98}_{71}	88^{99}_{72}
$J_{rem_{ps}}$	11^{29}_{1}	16^{29}_{2}	12^{29}_{1}	14^{29}_{2}	12^{28}_{1}	14^{29}_{2}	15^{29}_{2}	12^{28}_{1}

Notes: The table reports the fate for the N surplus: the stream export ($J_{out_{sub}}$), the total denitrification in source zone and subsurface ($J_{den_{tot}}$), the in-stream removal ($J_{rem_{sub}}$), the change in the source zone storage (ΔM_s) which includes three N stores (organic protected, organic active, and inorganic N stores), and the change in the subsurface storage (ΔM_{sub}). It also reports the fate for the N point sources (J_{ps}): the stream export ($J_{out_{ps}}$), and the in-stream removal ($J_{rem_{ps}}$). For simulated variables, numbers indicate the median, and lower bound (lb) and upper bound (ub) of the 95% CI in the behavioural simulation ensemble: $median^{ub}_{lb}$.

5.2.2 Contribution of the N Point Sources to the In-stream N Loading

We investigate the contribution of the in-stream N loading originating from N point sources ($J_{out_{ps}}$) to the total in-stream loading (J_{out}) over the period 1960–2015. For Hemelingen, we find that the N point sources are an important flux that amounts to 217 $kg\ ha^{-1}$, and that accounts for 28.7% of the total in-stream N loading (Table 2 and supplementary Table S13). For all subcatchments, the N points sources contribution to the total in-stream N loading is between 20% and 29%, apart from Verden for which it is as high as 34.4% (median values, as reported in supplementary Table S13). We note that the 95% CI on the point sources contribution is rather large, as e.g. for Hemelingen it is 22.6–35.6%. This can be partly explained by the large uncertainty in the in-stream N removal (Table 2), since behavioural estimates of R can span over its entire prior range (0.01-0.3) for all subcatchments (Figure 5). For the last ten years of the simulation period (2006–2015), i.e., when the point sources is at its lowest level, N point sources still contribute to between 14% and 20% of the total in-stream N loading (95% CI: 10–26%) across all subcatchmnents (supplementary Table S13).

To understand the relative role of point and diffuse (N surplus) sources on the resulting temporal trend of the total in-stream N concentration, we perform a piecewise linear trend analysis for each individual component of the concentration over different time periods for Hemelingen. The analysis is based on the median of the behavioural simulation ensemble (total concentration is represented by a dashed red line in Figure 4-h). For the period 1970–1990, with respect to the total concentration, we find no statistically significant trend (significance level: 0.01) and a very small slope of the regression line (s_{lin}=-0.01 $mg\ L^{-1}\ yr^{-1}$), which is explained by the contrasting trends in the diffuse sources contribution (negative slope of the regression line s_{lin}=-0.08 $mg\ L^{-1}\ yr^{-1}$) and point sources contributions (positive slope of the regression line s_{lin}=0.07 $mg\ L^{-1}\ yr^{-1}$). For the period 1990–2000, the total concentration shows a marked decreasing trend (s_{lin}=-0.16 $mg\ L^{-1}\ yr^{-1}$). Over the same time period, the point sources show a stronger decline (s_{lin}=-0.1 $mg\ L^{-1}\ yr^{-1}$) compared to the diffuse sources (s_{lin}=-0.05 $mg\ L^{-1}\ yr^{-1}$). During the last time period 2000–2015, the concentration trends are either non-significant for the total concentration and the point sources contribution, or small for the diffuse sources contribution (s_{lin}=-0.02 $mg\ L^{-1}\ yr^{-1}$). The concentration time series used for this trend analysis and the regression lines are reported in supplementary Figure S32.

5.3 Uncertainty and Sensitivity of the Simulated N Legacies

Section 5.2.1 mostly focuses on examining the median values of the simulated N legacies. However the uncertainty is large (Table 2), due to the limited information available to constrain these legacies. The soft rules hardly affect the distribution of the simulated change in source zone storage (ΔM_s, top panel of Figure 7). The width of the 95% CI in the behavioural ensemble is about equal to the median value, apart from Verden for which it is 1.7 time higher (red boxplots). Regarding the subsurface, the soft rules can have a contrasting effect on the simulated change in storage (ΔM_{sub}, bottom panel of Figure 7), as they can reduce but also exacerbate the uncertainty. Values that are outliers, i.e., beyond the 95% CI of the grey boxplots, in the unconstrained ensemble can be identified as behavioural and be included in the constrained 95% CI (colored boxplots). The width of the 95% CI for ΔM_{sub} is between two to three times higher than the median value.

Here, we investigate the factors that explain this residual uncertainty in the legacy N stores by assessing the sensitivity of the simulated N legacies to the ELEMeNT parameters, the N point sources and the N surplus in the constrained simulation ensemble obtained after application of the soft rules, for Hemelingen (see method in Section 4.2). For each of the 270 combinations of N surplus and N point sources, we identify between 531 and 2146 behavioural simulations (details in supplementary Table S14 and Figures S33–35). This results in a total sample of the model input-output of size 362,985 to perform the PAWN sensitivity analysis. We observe that the bootstrap confidence intervals of the estimated sensitivity indices are narrow and exhibit little overlap among the different inputs (Figure 8). Therefore, the sample size is sufficient to infer a robust ranking of the input factors according to their relative importance.

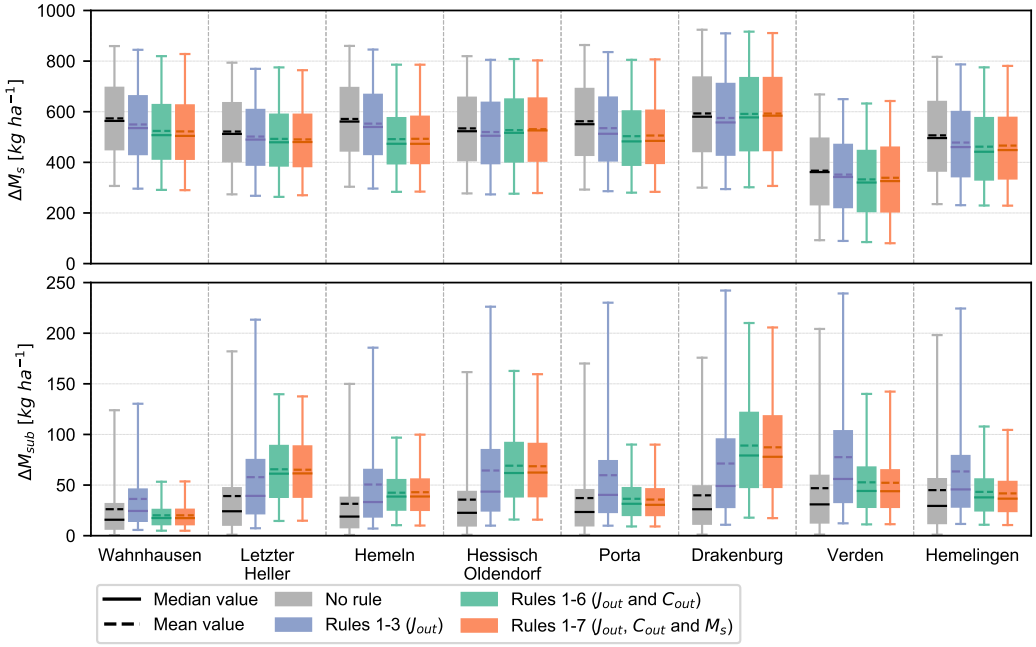

Figure 7. Statistics of the simulated N legacies for the period 1960–2015 (change in source zone N storage – ΔM_s, and change in subsurface N storage – ΔM_{sub}) obtained for different levels of constraining. The source zone N storage includes the three N storages (organic protected, organic active, and inorganic N stores). The boxplots report the median (solid line), mean (dashed line), 2.5%, 25%, 75% and 97.5% quantiles of the simulation ensembles.

From Figure 8 we observe that the N point sources (PS) and two parameters used to generate the N surplus realizations, namely the ratio of the N surplus for agricultural permanent grassland to the N surplus for cropland ($r_{mgra-crop}$) and the ratio of the agricultural N surplus in 1850 to the value in 1950 (r_{warm}) have very small sensitivity indices and are the least sensitive inputs for all four output variables considered. This means that the uncertainties in the ELEMeNT parameters and in the value of the total N surplus (multiplier parameter $f_{surplus}$) have a much larger impact on the behavioural values of the legacy stores than the uncertainties in PS, $r_{mgra-crop}$, and r_{warm}.

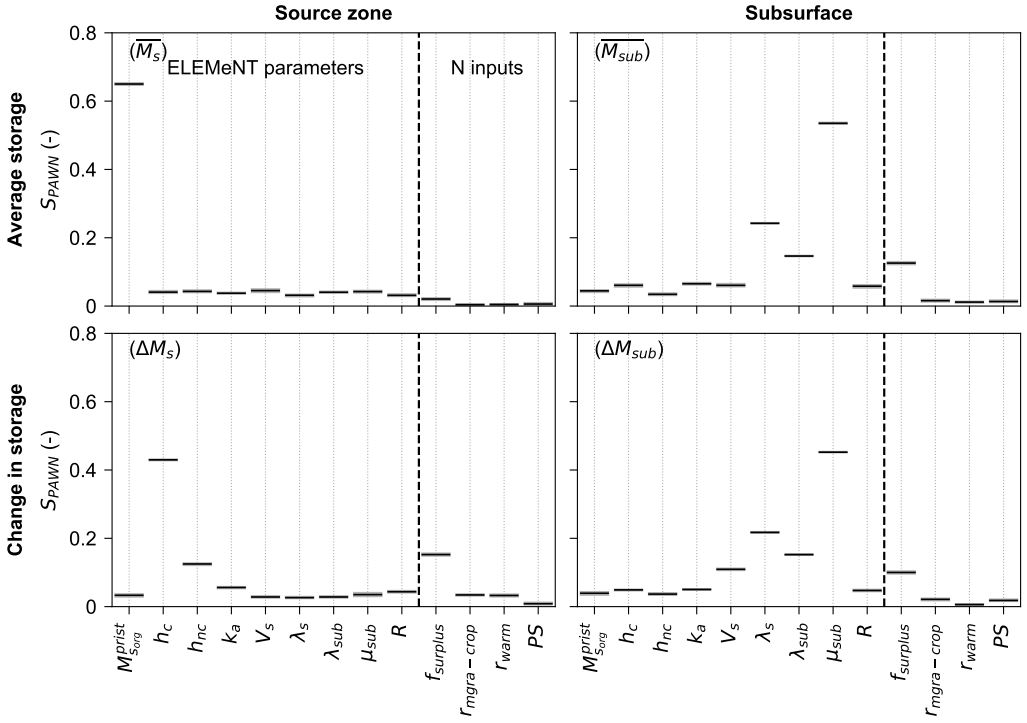

Figure 8. PAWN sensitivity indices S_{PAWN} of the nine ELEMeNT parameters, the three parameters introduced to generate alternative N surplus realizations ($f_{surplus}$, $r_{mgra-crop}$, and r_{warm}), and the N point sources realization (PS), for the WRB at Hemelingen. Sensitivity indices are reported with respect to four model outputs evaluated over the period 1960–2015, namely the average source zone N storage $\overline{M_s}$, the average subsurface N storage $\overline{M_{sub}}$, the cumulative change in source zone N storage ΔM_s, and the cumulative change in subsurface N storage ΔM_{sub}. The source zone N storage includes the three N storages (organic protected, organic active, and inorganic N stores). The horizontal black lines indicate the bootstrap mean value of the sensitivity indices, while the grey boxes represent the 95% bootstrap confidence intervals. The bootstrap confidence intervals are very small (the grey boxes are very narrow), since the size of the sample used to calculate the PAWN indices is very large.

For the source zone, the sensitivity analysis results with respect to the average N storage ($\overline{M_s}$) and the change in N storage (ΔM_s) differ. The protection coefficient for cultivated land (h_c) is largely responsible for the uncertainty in ΔM_s, followed by the N surplus multiplier ($f_{surplus}$) and the protection coefficient for non-cultivated land (h_{nc}) whose sensitivity indices have similar magnitude. In contrast, the source zone organic N stock under pristine conditions ($M_{s_{org}}^{prist}$) is by far the most influential parameter for $\overline{M_s}$. For the subsurface zone, unlike the source zone, results are similar for the two statistics analyzed (the average N storage $\overline{M_{sub}}$ and the change in N storage ΔM_{sub}), and a larger number of input factors are influential. Specifically, the mean travel time in the subsurface (μ_{sub}) is the most influential input, followed in decreasing order of importance by the denitrification rate constants in the source zone (λ_s) and in the subsurface (λ_{sub}), and the N surplus multiplier ($f_{surplus}$). In addition the mean annual water content in the source zone (V_s) has a stronger impact on the change in the subsurface N storage than on the average storage. The value of the sensitivity index of V_s with respect to the change in subsurface storage is similar to the sensitivity index of $f_{surplus}$.

The importance of the three N surplus parameters and the N point sources may be higher than suggested by the PAWN analysis because the parameter estimation may result in different posterior parameter distributions when using different N input realizations. Therefore, the application of the soft rules may compensate for the uncertainties in the N inputs. This effect is particularly visible for the N surplus multiplier, which has some impact on the distributions of the mean travel time and the denitrification rate constants (supplementary Figure S38), but it is much less pronounced for the other N inputs parameters (supplementary Figure S36–37).

We further examine the robustness of the PAWN analysis by estimating the sensitivity indices using other summary statistics (mean and maximum values) than the median value, to aggregate the KS values across the conditioning intervals in Equation 10. As shown in Figures S39–S40, the results show a similar order of importance among all inputs (model parameters and N input realizations) as shown in Figure 8. We however note that, when using the maximum KS, the mineralization rate constant for the active store (k_a) has a sensitivity index of the same magnitude as $f_{surplus}$ with respect to the average subsurface storage. From supplementary Figures S41–S44 that report the conditional and unconditional CDFs used for the calculation of the PAWN indices, we see that k_a has a higher impact in the lower 10% of its range (values lower than 0.25 yr^{-1}) on the average subsurface storage.

6 Discussion

6.1 Performance of the Simulated In-stream N Loading and Concentration

The ELEMeNT model is able to produce simulations that are consistent with observations of in-stream N loading and concentration (i.e., that show satisfactory values for each of the three components of the KGE), and with the source zone N content in 2009 for the different study catchments (Figures 2-4). While previous studies using the ELEMeNT model (Chang et al., 2021; J. Liu et al., 2021; Van Meter et al., 2017) focused on simulating the in-stream N loading, here we also examine the concentration, which is more difficult to simulate than the loading (Husic et al., 2019). As seen in our results, the loading dynamics are predominantly determined by the discharge dynamics, while the concentration dynamics may be the results of more complex processes. It is important to assess both N loading and concentration to characterize the in-stream water quality status, as N loading affects the status in downstream receiving water, while concentration describes the local water quality status (Hirsch et al., 2010).

Although, our simulations are overall in good agreement with the observations, we observe a discrepancy between the median simulated N concentration and the measured values for the later years (2005–2015). This is particularly visible at Verden, where the concentration observations are at the lower end or lower than the behavioural simulation ensemble for this time period. First, this mismatch between simulated and observed concentration could be due to changes in the characteristics of the landscape (such as e.g. the density of tile drains as observed in J. Liu et al. (2021)), that could require temporally varying parameter values. However, no information is available to us to substantiate that such changes have occurred. Second, another possible cause for the discrepancy between observed and simulated concentration could be the uncertainty in the N input data (N surplus and N point sources). In particular, in Germany, data on application of mineral fertilizer exist at the national level only. For lower administrative levels, data refer to the sale of mineral fertilizer and are therefore strongly linked to the location of the fertilizer companies, rather than the actual amount of fertilizer application. Therefore, subjective choices have to be made to disaggregate the fertilizer application amounts to finer spatial units (Häußermann et al., 2020; Behrendt et al., 2003). Fertilizer application is an important component of the N surplus and changes have been implemented after 2007 due to a new ordinance that limits fertilizer use (DüV, 2007). This could contribute to the uncertainty in our in-stream

simulations. In Section 6.3.2, we further elaborate on the need for a better estimation of the uncertainty in the N surplus for an improved characterization of not only the in-stream variables, but also the N legacy stores.

6.2 N Mass Balance in the WRB and its Implications for Water and Land Management

6.2.1 Denitrification in the Terrestrial Compartments

Our results (Figure 6 and Table 2) indicate that denitrification in the terrestrial system (source zone and subsurface) is the largest sink for the N surplus in the WRB for the period 1960–2015. This is consistent with previous modelling studies of N legacies performed in North America, namely Ilampooranan et al. (2019) using the SWAT-LAG model, and J. Liu et al. (2021) and Van Meter et al. (2017) using the ELEMeNT model. However, we find that denitrification is much higher for the WRB, in that it is likely to be higher than 50% for the eight subcatchments, while it is found to be less than 50% in the three previous studies. Such a high amount of denitrification in the WRB could have adverse consequences on the atmosphere and the climate, because it can potentially release nitrous oxide (N_2O). Yet, N_2O can be further reduced into harmless dinitrogen N_2 in presence of favourable environmental conditions in the soil and subsurface (Betlach & Tiedje, 1981; Bergsma et al., 2002; Robertson & Groffman, 2015; Rivett et al., 2008). The results of ongoing efforts, such as the "Global N_2O Model Intercomparison Project" (Tian et al., 2018), will be key to better understand and quantify the processes involved in N_2O emissions and the impact of denitrification on the atmosphere.

6.2.2 Accumulation of N Legacies

In this study, we explicitly quantify the N legacies in the WRB over the period 1960–2015 (Figure 6 and Table 2). Previous N modelling studies in the WRB are either based on a simple regression model that does not account for N storages in the terrestrial system (GREEN model; Grizzetti et al., 2008)), or on a modelling framework that represents the residence time of N in the groundwater, but that does not consider source zone N storage nor assesses the long-term accumulation of N in the subsurface (Heidecke et al., 2015; Hirt et al., 2012; Kreins et al., 2010). We find that the N accumulation in the source zone, which amounts to 15% of the N surplus (95% CI: 8–26%), is an order of magnitude higher that

the N accumulation in the subsurface, which is equal to less than 3% of the N surplus. The magnitude of the simulated N accumulation in the source zone and the subsurface is similar across subcatchments. Groundwater nitrate concentration is found to be higher than the regulatory threshold of 11.3 $mg\ L^{-1}$ in some measuring points in the WRB, in particular in Lower Saxony (NLWKN, 2019), while background levels are typically very low. This supports our result that N has been building up in the subsurface. However, no information is available on the N accumulation in the source zone to corroborate our findings. In previous studies of N legacies, the relative value of biogeochemical and hydrologic N legacies is highly variable. Whereas the N buildup in the subsurface store is also found to be lower than the buildup in the source zone in the Grand river basin (J. Liu et al., 2021), a subcatchment of the Iowa-Cedar basin (Ilampooranan et al., 2019) and the Mississippi river basin (Van Meter et al., 2017), the opposite result is reported for the Susquehanna river basin (Van Meter et al., 2017). The magnitude of hydrologic N legacies in J. Liu et al. (2021) is similar to our study (around 4% of the N surplus), while it reaches 14% in Ilampooranan et al. (2019).

Although the N accumulation in the subsurface represents a small fraction of the N surplus in the WRB, this N store is composed of reactive and dissolved N forms, which can be easily accessed and mobilized. Thus, they are of immediate relevance for the water quality status. Since the mean travel time in the subsurface is found to be equal to 8 yr (95% CI: 2–24 yr), the subsurface N storage is likely to impact the stream N concentration over the coming years. In the source zone, the accumulated N could be a threat for future water N levels as well, depending on how fast it can mineralize. Findings of previous studies suggest that applied N fertilizer can slowly leach over decades following their application (Haag & Kaupenjohann, 2001; Sebilo et al., 2013). Our results indicate that most of the N accumulation occurs in the organic protected N pool (median value for Hemelingen: 94%, Table S12), whose transfer time is in the order of magnitude of a few millennia (Table S9). The N buildup is much smaller in the active pool that has a transfer time of a few years to a decade. The long transfer times of N stored in the source zone, which can be made possible through the protection of N into organic matter (Six et al., 2002), may partly explain why it is difficult to lower the N concentrations to acceptable values in the WRB and, more broadly, in regions with long history of high N inputs, as described e.g. in Grimvall et al. (2000) and Vero et al. (2018). Source zone N storage could also be a potential resource for crop growth, allowing satisfaction of the crop N requirements with lower amounts of fertilizer application, as proposed by Dupas et al. (2020) and J. Liu et al. (2021). In particular, while

it is widely accepted that crops can use mineral N compounds present in the soil, they may also take up organic N forms, but this process is still not well understood (Näsholm et al., 2009; Farzadfar et al., 2021). Therefore, the fate of the N stored in the source zone has large uncertainties and depends on the ability of the plants to access it and on its potential to mineralize to yield more available N forms.

Regarding the temporal dynamics, the permanent buildup in the source zone found for the WRB is consistent with most of the previous N legacy studies (Ilampooranan et al., 2019; J. Liu et al., 2021; Van Meter et al., 2016). The WRB shows a large decrease in the subsurface N store during the period 1990–2010, that can be explained by the concurrent reduction in the N surplus. Hydrologic N legacies permit to sustain higher in-stream concentration levels over this time period. This is particularly visible for the Letzter Heller subcatchment, which has undergone a large and sudden decrease in the N surplus after 1990, while in-stream N concentration remains relatively stable. This result for the subsurface N legacy differs from earlier legacy studies over the Mississippi and Susquehanna river basins and the subcatchment of the Iowa-Cedar river basins, where the subsurface N store is continuously building up in time (Ilampooranan et al., 2019; Van Meter et al., 2016). Given the importance of the N legacies for land and water management and in particular to achieve the target of 2.8 mg L^{-1} for in-stream concentration in the WRB (OGewV, 2016), better characterization and reduction of the uncertainties in the simulated N legacies is crucial, as further discussed in Section 6.3.

6.2.3 Importance of the N Point Sources

We find that N point sources from wastewater represent an important fraction of the in-stream nitrate loading in the WRB (Figure 6, Tables 2 and S13). Point sources N loads comprise 28.7% (95% CI: 22.6–35.6%) of the stream N load for the period 1960–2015. The contribution is smaller for the later period (2006–2015), where point sources have a lower magnitude due widespread connection to wastewater treatment plants and high efficiency of treatment. Grizzetti et al. (2008) find that point sources account for 31% of the stream N-NO_3 loading over the period 1995–2002 for the WRB at Hemelingen. This is higher than our uncertainty estimates for this time period (95% CI: 13–23%, Table S13), which could be explained by the differences in the model structure used in Grizzetti et al. (2008) (regression based GREEN model) and in the N point sources inputs. In our study, the N point sources are constrained by recent observations of N loading from wastewater treatment

plants (Section 3.2.4), while Grizzetti et al. (2008) do not make use of observational data. Moreover, the temporal variations in the N point sources have a large effect on the trend of the total in-stream N concentration during the period 1970–2000 (Figure S32). The decrease in the N point sources during the 1970s and 1980s counteracts the increase in the contribution of the N diffuse sources (N surplus) to the in-stream N concentration, resulting in no overall trend in the total concentration. The marked decrease in the total concentration in the 1990s is also largely dominated by the decrease in the point sources. While the N diffuse sources are the largest contributor in magnitude to the in-stream concentration, their temporal signal can be smoothed through biogeochemical transformations and transport in the source zone and subsurface (Figure S32). In contrast, changes in the N point sources have an immediate impact on the in-stream concentration and can therefore strongly influence its trend.

Some past N modelling studies covering a large range of catchments across Germany and France have not accounted for N point sources (Dupas et al., 2020; Ehrhardt et al., 2021). Based on our simulation results, we recommend the consideration of N point sources and their temporal variations in future N modeling analyses over the WRB.

6.3 Towards Reducing the Uncertainty and Equifinality in the Simulations of a N Model

6.3.1 Value of the Soft Rules to Constrain the Model Uncertainties

In this study, we utilize three different sets of observational data (in-stream N loading and concentration and source zone N content) to estimate the model parameters, using a portfolio of soft rules to constrain the model results. We show that, beyond in-stream N loading, in-stream concentration and source zone N content have a value in reducing the number of behavioural simulations and in constraining the parameter distributions, thus reducing the equifinality (Figure 2b, Figure 5). Specifically, the in-stream N loading and concentration data affect the simulated in-stream loading and concentration (Figures 3,4) and the simulated change in the subsurface storage (Figure 7). The source zone N content is the only data that allows to constrain the magnitude of the total simulated N storage (Figure S17), but it has no appreciable impact on the different components of the mass balance (Figures 3,4,7). Importantly, the soft rules do not constrain the change in the source zone N store (Figure 7).

Only few previous N modelling studies analyzed the equifinality by performing a detailed investigation of the parameter space, including Husic et al. (2019) and Rankinen et al. (2006). Due to the different model structures used in these studies, our parameter estimation results cannot be directly compared to these studies. Yet, we note that Rankinen et al. (2006) reveals a strong interplay between terrestrial and in-stream model processes in a subcatchments of the Simojokiriver basin in Finland. This is consistent with our results, as we could barely constrain the in-stream N removal parameter (R).

Despite the equifinality we can constrain the range of a few parameters, including the mean travel time in the subsurface μ_{sub}, which we determine to be equal to 8 yr (95% CI: 2–24 yr) at Hemelingen (Table S7). Koeniger et al. (2008) reports values of the mean groundwater travel time in the range 8-93 yr (37 yr for total flow) for Hemelingen, based on long-term tritium isotope data in combination with simulations from a hydrologic model. Hirt et al. (2012) found a value of the mean travel time in the groundwater of 25 yr. Ehrhardt et al. (2021) established that the overall mode travel time for total flow are in the range 0-20 yr for different subcatchments of the WRB and in particular in the range 0-10 yr for most subcatchments. Therefore previous studies cover a large spectrum of travel time values, which includes our estimates. The results of the study of Hirt et al. (2012) are also, as expected, on the higher end of our uncertainty estimates. This can be because Hirt et al. (2012) explicitly account for quicker flow paths (tile drainage) in their modelling framework, while in ELEMeNT all flow paths to the stream are lumped in the subsurface compartment.

The soft rules only allow to reduce part of the equifinality, as some parameter distributions could be hardly constrained (Figure 5) and the uncertainty in the model internal components remains large (Figure 7). For example, it would be relevant to determine the amount of denitrification occurring in the source zone and the subsurface, because denitrification in the subsurface can involve the irreversible degradation of substances, such as pyrite, which is not sustainable (Wendland et al., 2009; Wriedt & Rode, 2006). Such quantification is however not possible due to equifinality (Figure 5, supplementary Table S11) and therefore this topic deserves further investigations. In addition, establishing a robust ranking of importance of the N legacy buildup between subcatchments would be desirable to target management efforts to legacy hotspots (J. Liu et al., 2021). However, the residual uncertainty in the simulated N legacies is still large (Figure 7) and the confidence intervals of the distributions of the N buildup obtained for the different subcatchments are overlapping.

This equifinality can be due to parameter interactions (Table S10) or to the fact that some model parameters are not influential with respect to the metrics used in the soft rules.

6.3.2 Strategies to Further Reduce the Uncertainty and Equifinality

To tackle this issue of uncertainty and equifinality, we perform a sensitivity analysis to investigate the factors that are responsible for the residual uncertainty in the simulated N legacy stores and that should be the focus of future efforts for uncertainty reduction (Figure 8). We apply the PAWN method, which does not rely on any assumption regarding the model input-output relationship, to the constrained input-output sample. The studies of Van Meter et al. (2017) and J. Liu et al. (2021) assess the sensitivity of the median source zone N store simulated with the ELEMeNT model. Although the method adopted in these two studies is linear regression and the sensitivity analysis is carried out based on the unconstrained sample before calibration, their results are comparable to our study. We observe that the most influential factor is by far the source zone organic N stock under pristine conditions ($M_{s_{org}}^{prist}$) in both our study and in J. Liu et al. (2021), and the mineralization rate constant for the organic protected pool, which is related to $M_{s_{org}}^{prist}$ (Equation S24) in Van Meter et al. (2017). We further note that Van Meter et al. (2017) and J. Liu et al. (2021) examine the sensitivity of the cumulative in-stream N loading and find, similar to our results for the average subsurface N store, that the mean travel time and the two denitrification rate constants in the source zone and the subsurface are the three most influential parameters. This similarity between the sensitivity of the in-stream N loading and the subsurface N store can be explained by the structure of the ELEMeNT model which lumps all flow paths to the stream into the subsurface store.

Our sensitivity analysis (Figure 8) reveals that the protection coefficient for cultivated land is mostly responsible for the residual uncertainty in the simulated accumulation in the source zone N store. This parameter could scarcely be constrained by the soft rules (Figure 5). The protection coefficient is a conceptual parameter that partitions the N surplus between the organic active and protected N stores, and therefore it can hardly be inferred through field measurements. A possible solution could be to further refine and constrain the representation of the protection mechanism in the source zone using information gained from simulation experiments carried out with more complex models, that focus on the soil processes and that can include a large number of soil N pools, such as the DAISY (Hansen et al., 1991) or the CANDY model (Franko et al., 1995) (a review of soil

organic matter models is provided in Campbell & Paustian, 2015). Regarding the buildup of the subsurface N store, three interacting parameters mostly contribute to the uncertainty, namely the mean travel time in the subsurface, which is also the most influential factor, and the two denitrification rate constants. In this regard, tracer studies, and in particular the combination of tritium concentration and helium isotope measurements, can help to characterize the travel time (Sültenfuß et al., 2009), as well as the modelling of conservative solutes, such as chloride, in combination with nitrate (Kaandorp et al., 2021). In addition, Eschenbach et al. (2018) propose a method to characterize denitrification in the groundwater, based on the measurement of the N_2/Ar ratio. Such techniques provide promising avenues for constraining denitrification fluxes and thereby possibly reducing the uncertainty in the simulated N legacies.

Regarding the magnitude of the total N surplus, we characterize the uncertainty of this input data by using a time-invariant multiplier coefficient and we explore a variation of $\pm 20\%$ with respect to the baseline N surplus data. N surplus datasets for Germany do not provide uncertainty intervals, as uncertainty estimation is currently not a common practice in N surplus construction (X. Zhang et al., 2021). An improved assessment of this uncertainty in future studies seem necessary, since, on the one hand, our results show that the uncertainty in the N surplus has an impact (1) on the simulated N legacies (Figure 8), although this impact is smaller than the ELEMeNT parameters that we discussed previously in this section, (2) on the posterior distribution of the model parameters (Figure S38), and (3) possibly on the simulated in-stream concentration trend over the period 2005-2015 (as discussed in Section 6.1). On the other hand, X. Zhang et al. (2021) found large discrepancies in different components of the N surplus for agricultural areas between different global datasets, which suggests that the actual uncertainties in the N surplus can be large.

We recognize that, in this study, the uncertainties on the different components of the N mass balance, including the simulated N legacy buildup, could be underestimated. In fact, we examine the uncertainties in the N surplus using exploratory coefficients that may not reflect the actual uncertainties in the N surplus. We also do not investigate the model structural uncertainties. To further address the modelling uncertainties and equifinality, future studies need to reveal not only the uncertainties in the model parameter values, but also in the data used as input or to constrain the simulations, and in the model structures, as elaborated below.

First, regarding the parameters values, the definition of the parameter distributions and ranges can affect the parameter estimation results, because they can greatly impact the sensitivity of the metrics used for calibration (e.g. the bias, variability error or correlation used in this study) to the model parameters (a discussion on the impact of the parameter ranges on sensitivity analysis results is provided e.g. in Pianosi et al., 2016). To address this issue, in this study, we define the ranges through a careful literature review (Table 1).

Second, with respect to the data uncertainties, in this study we focus on the N inputs and the soil N data (used in the soft rules). Including uncertainty on soil N data is crucial especially, if the objective is to detect a change in storage when data are provided for different years. Given the large size of the total N storage (Table S12), the change in storage may be within the observation uncertainties. In our study, we consider that other data uncertainties are smaller. In fact, the annual stream discharge data is the combination of observations, and evaluated and bias-corrected model simulations (Section 3.2.5). In contrast, no observational data of N surplus exist to test the plausibility of the N surplus estimates. We also consider that the in-stream N concentration data has a high quality, since they come from 14-day average measurements (Section 3.2.6). However, uncertainty should be examined when data have a lower quality, in particular when using low-frequency in-stream concentration observations combined with weighted regressions on time, discharge and season (WRTDS, Hirsch et al., 2010). In this respect, a bootstrap approach could be envisaged (Hirsch et al., 2015).

Third, a modelling approach allowing for systematic exploration of different modelling alternatives could be developed, similar to the Structure for Unifying Multiple Modeling Alternatives (SUMMA, Clark et al., 2015a, 2015b) that allow testing of alternative model formulations for a range of different hydrological and thermodynamic processes. Specifically, in the source zone, worth of further investigation are the representation of the processes of immobilization of N into organic matter and of N saturation, which are poorly characterized (Bingham & Cotrufo, 2016; Yansheng et al., 2020). In the subsurface, further mixing schemes beyond complete mixing/random sampling could be examined using StorAge Selection (SAS) functions, as implemented e.g. in Nguyen et al. (2021).

To help identification of plausible model structures and parameters values, our study call for the long-term monitoring of N content in the soil and along the subsurface (unsaturated zone and groundwater) profile. Current N data in the subsurface are typically provided

at a unique depth below the water table at each measuring site, e.g. in the European Waterbase dataset (EEA, 2021) or in the dataset provided by the German state of Lower Saxony (NLWKN, 2022). These data do not allow straightforward quantification of the subsurface N storage, since nitrate concentration can vary greatly with the depth to the water table (MacDonald et al., 2017; Rudolph et al., 1998) and large amounts of N could also be stored in the unsaturated zone (Ascott et al., 2017). Regarding the soil part, future modelling studies could make use of the data on soil mineral N content that will likely become available, in particular in Germany where the 2017 fertilizer ordinance (DüV, 2017) prescribes the investigation of soil mineral N prior to fertilizer application.

Yet, due to the scale mismatch between point scale measurements of soil and subsurface N content and the modelling resolutions, the incorporation of these data into the modelling exercise requires the use of smart techniques and appropriate model structures that are commensurate with the measurements (Peters-Lidard et al., 2017). Moving towards the use of spatially distributed water quality models (X. Yang et al., 2018; Nguyen et al., 2021) may be a way forward for integrating local scale measurements into the modelling framework. Such models should be however combined with smart parameterization techniques, such as the Multiscale Parameter Regionalization (MPR, Kumar et al., 2013; Samaniego et al., 2010), which allows for seamless simulations at multiple scales and facilitates the incorporation of finer level information (Rakovec et al., 2016; Samaniego et al., 2017).

7 Conclusions

The objectives of this study were to 1) characterize the uncertainties in the long-term fate of the N inputs to the landscape, simulated with a parsimonious catchment-scale N model (ELEMeNT), 2) determine the value of different (observational) data to constrain the simulation results with emphasis on the simulated N legacies, and 3) gain further understanding of the magnitude and dynamics of the N legacies to determine their relevance for water and land management. To do so, we establish the ELEMeNT model in eight nested sub-catchments of the WRB, and simulate the fate of N and the dynamics of the legacy stores over the last six decades (1960–2015).

We introduce a multicriteria parameter estimation strategy based on soft rules, that imposes acceptability limits to the model performance in reproducing the in-stream N loading and concentration, and the source zone N content in 2009. We demonstrate that this

procedure allows to reduce the equifinality. In particular, the in-stream data allow to constrain the simulated in-stream N loading and concentration and the change in the subsurface N storage, while the source zone N content data reduce the uncertainty in the simulated total source zone N storage. All sources of information also have value in constraining the parameter distributions. However, despite the parsimonious structure of the ELEMeNT model, the uncertainties in the mass balance components remain substantial after using all available information to constrain the simulations. This is due to equifinality, and more specifically to interactions between the model parameters, e.g. between the travel time in the subsurface and the denitrification rate constants. Our sensitivity analysis reveals crucial information on model functioning by identifying key model parameters, such as the protection coefficient for cultivated land, the travel time in the subsurface and the dentrification rate constants in the source zone and the subsurface, that are largely responsible for the residual uncertainty in the simulated N legacies. The N surplus input could also be an important source of uncertainty. It uncertainty estimates should be better assessed in future works to refine the exploratory multiplier coefficient approach used in this study.

Given our modelling assumptions and the data we used, our simulation results suggest a relative importance of the different constituents of the N mass balance in the WRB over the period 1960–2015. Denitrification is found to be the largest sink for the N surplus and is likely to be higher than 50%, followed by the in-stream N export and source zone N accumulation – both with similar magnitude (median value: 15–18%), while subsurface N accumulation and in-stream N removal appear to be the smaller components (lower than 4%). Total accumulation in legacies stores in the WRB amounts to around 491 $kg\ ha^{-1}$ (95% CI: 264–820 $kg\ ha^{-1}$). Although the buildup of the subsurface N store represents a small proportion of the N surplus, it constitutes an immediate threat for the water quality status, since it includes mobile N forms. Furthermore, our analysis reveals N point sources as one of the important contributor to the in-stream N levels (median value: 28.7% over the period 1960–2015); and therefore we recommend that more attention should be given to this component to properly analyze N dynamics in future modeling studies.

Overall, we recognize that our simulation results have large uncertainties. Our study calls for a thorough consideration of equifinality in catchment water quality modelling, for a better characterization of the model internal components, such as the biogeochemical and hydrologic N legacies. Although knowledge about N legacies is crucial to reach the water quality goals and improve the ecological status of water bodies, this topic deserves more

attention. In particular, modelling of N legacies is fraught with a myriad of uncertainties arising from different sources, including not only the model parameter values and input data that are examined in this study, but also the model structures and sparse measurements (e.g. low-frequency in-stream concentration observations). To this end, we believe that sensitivity analysis can be a promising tool for tackling the uncertainty and equifinality. In fact, it allows identification and pinpointing of the model input factors that are responsible for the uncertainty and that should be the focus of future efforts for uncertainty reduction. Importantly, spatially lumped or semi-distributed model structures may restrict the amount of observational data that can be incorporated into the modelling framework, because of the incommensurability between the data, and the model parameter and corresponding simulations. Therefore, future efforts towards reducing the equifinality should focus on both collecting further data, and improving the model representations (e.g. parameterization and structures).

Open Research

The land use, N surplus, N point sources and mHM simulated discharge data, as well as the ELEMeNT simulated N output as available at `https://www.hydroshare.org/resource/8779a09b9f204172931a641dd27d00c4/`. The underlying data used in this study were downloaded from: `https://datenbank.fgg-weser.de/weserdatenbank` (in-stream nitrate concentration and river discharge data, FGG Weser, 2021); `http://www.bafg.de/GRDC` (river discharge data Global Runoff Data Centre, 2021); `https://dataportaal.pbl.nl/downloads/HYDE/` (History Database of the Global Environment – HYDE, Klein Goldewijk et al., 2017); `https://esgf-node.llnl.gov/` (atmospheric N deposition data); `https://land.copernicus.eu/pan-european/corine-land-cover` (Corine Land Cover, EEA, 2019a); `https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data` (topsoil bulk density based on the Land Use and Cover Area frame statistical Survey – LUCAS data, Ballabio et al., 2016); `https://esdac.jrc.ec.europa.eu/content/chemical-properties-european-scale-based-lucas-topsoil-data` (soil N content based on LUCAS data, Ballabio et al., 2019); `http://www.fao.org/faostat/en/\#data/FBSH` and `http://www.fao.org/faostat/en/\#data/FBS` (protein supply data, FAO, 2021a, 2021b); `http://appsso.eurostat.ec.europa.eu/nui/show.do?lang=en\&dataset=env\_ww\_con` and `https://db.nomics.world/Eurostat/env\_wwcon\_r2` (population connection to sewer and wastewater treatment, Eurostat, 2016, 2021).

Acknowledgments

Support to FS was provided by the Reduced Complexity Models project co-funded by the Helmholtz Association. FS, RK, MW and SA acknowledge the Advanced Earth Modelling Capacity (ESM) project funded by the Helmholtz Association. We thank Martin Bach for providing the N surplus data for agricultural areas. We thank Oldrich Rakovec for providing the long-term mHM simulations. We thank two anonymous referees for their useful suggestions that allowed us to improve the quality of this paper.

References

Arle, J., Blondzik, K., Claussen, U., Duffek, A., Grimm, S., Hilliges, F., ... Wolter, R. (2017). *Waters in Germany: Status and assessment German environment.* Dessau-Roßlau, Germany: German Environment Agency.

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*, 1416. doi: 10.1038/s41467-017-01321-w

Bach, M., Breuer, L., Frede, H., Huisman, J., Otte, A., & Waldhardt, R. (2006). Accuracy and congruency of three different digital land-use maps. *Landscape and Urban Planning*, *78*(4), 289–299. doi: 10.1016/j.landurbplan.2005.09.004

Bach, M., & Frede, H.-G. (1998). Agricultural nitrogen, phosphorus and potassium balances in Germany — Methodology and trends 1970 to 1995. *Zeitschrift für Pflanzenernährung und Bodenkunde*, *161*(4), 385–393. doi: 10.1002/jpln.1998.3581610406

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., ... Panagos, P. (2019). Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma*, *355*, 113912. doi: 10.1016/j.geoderma.2019.113912

Ballabio, C., Panagos, P., & Monatanarella, L. (2016). Mapping topsoil physical properties at European scale using the LUCAS database. *Geoderma*, *261*, 110–123. doi: 10.1016/j.geoderma.2015.07.006

Basu, N. B., Rao, P. S. C., Thompson, S. E., Loukinova, N. V., Donner, S. D., Ye, S., & Sivapalan, M. (2011). Spatiotemporal averaging of in-stream solute removal dynamics. *Water Resources Research*, *47*(10), W00J06. doi: 10.1029/2010WR010196

Batjes, N. (1996). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, *47*(2), 151–163. doi: 10.1111/j.1365-2389.1996.tb01386.x

Behrendt, H., Bach, M., Kunkel, R., Opitz, D., Pagenkopf, W.-G., Scholz, G., & Wendland, F. (2003). *Nutrient Emissions into River Basins of Germany on the Basis of a Harmonized Procedure (Research report 299 22 285)*. Berlin, Germany: German Federal Environment Agency (Umweltbundesamt). Retrieved from `https://www.umweltbundesamt.de/en/publikationen/nutrient-emissions-into-river-basins-of-germany-on` (last access: 07 November 2021)

Behrendt, H., Huber, P., Kornmilch, M., Opitz, D., Schmoll, O., Scholz, G., ... Schweikart, U. (2000). *Nutrient emissions into River Basins of Germany (Research report 296 25 515)*. Berlin, Germany: German Federal Environment Agency (Umweltbundesamt). Retrieved from `https://www.umweltbundesamt.de/en/publikationen/nutrient-emissions-into-river-basins-of-germany` (last access: 07 November 2021)

Benettin, P., Rinaldo, A., & Botter, G. (2015). Tracking residence times in hydrological systems: Forward and backward formulations. *Hydrological Processes*, *29*(25), 5203–5213. doi: 10.1002/hyp.10513

Bergsma, T. T., Robertson, G. P., & Ostrom, N. E. (2002). Influence of soil moisture and land use history on denitrification end-products. *Journal of Environmental Quality*, *31*(3), 711–717. doi: 10.2134/jeq2002.7110

Betlach, M. R., & Tiedje, J. M. (1981). Kinetic explanation for accumulation of nitrite, nitric oxide, and nitrous oxide during bacterial denitrification. *Applied and environmental microbiology*, *42*(6), 1074–1084. doi: 10.1128/AEM.42.6.1074-1084.1981

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1-2), 18–36. doi: 10.1016/j.jhydrol.2005.07.007

Bingham, A. H., & Cotrufo, M. F. (2016). Organic nitrogen storage in mineral soil: Implications for policy and management. *Science of The Total Environment*, *551-552*, 116–126. doi: 10.1016/j.scitotenv.2016.02.020

Blankenau, K., Olfs, H.-W., & Kuhlmann, H. (2001). Immobilization of fertilizer nitrogen and consequences for nitrogen fertilization of cereals. In W. J. Horst et al. (Ed.), *Plant Nutrition. Developments in Plant and Soil Sciences* (Vol. 92, pp. 796–797). Dordrecht: Springer. doi: 10.1007/0-306-47624-X_387

BMEL. (2014). *Der Wald in Deutschland: Ausgewählte Ergebnisse der dritten Bundeswaldinventur*. Berlin, Germany: Bundesministerium für Ernährung und Landwirtschaft. Retrieved from `https://www.bmel.de/SharedDocs/Downloads/DE/`

Broschueren/bundeswaldinventur3.pdf?__blob=publicationFile&v=3 (last access: 25 January 2022)

Botter, G., Bertuzzo, E., & Rinaldo, A. (2010). Transport in the hydrologic response: Travel time distributions, soil moisture dynamics, and the old water paradox. *Water Resources Research*, *46*(3), W03514. doi: 10.1029/2009WR008371

Büttner, O. (2020). *DE-WWTP - data collection of wastewater treatment plants of Germany (status 2015, metadata).* HydroShare [data set]. doi: 10.4211/hs.712c1df62aca4ef29688242eeab7940c

Byrnes, D. K., Van Meter, K. J., & Basu, N. B. (2020). Long-Term shifts in U.S. nitrogen sources and sinks revealed by the new TREND-nitrogen data set (1930–2017). *Global Biogeochemical Cycles*, *34*(9), e2020GB006626. doi: 10.1029/2020GB006626

Campbell, E. E., & Paustian, K. (2015). Current developments in soil organic matter modeling and the expansion of model applications: a review. *Environmental Research Letters*, *10*(12), 123004. doi: 10.1088/1748-9326/10/12/123004

Castellano, M. J., Kaye, J. P., Lin, H., & Schmidt, J. P. (2012). Linking carbon saturation concepts to nitrogen saturation and retention. *Ecosystems*, *15*, 175—187. doi: 10.1007/s10021-011-9501-3

Chang, S. Y., Zhang, Q., Byrnes, D. K., Basu, N. B., & Meter, K. J. V. (2021). Chesapeake legacies: the importance of legacy nitrogen to improving Chesapeake Bay water quality. *Environmental Research Letters*, *16*(8), 085002. doi: 10.1088/1748-9326/ac0d7b

Chen, D., Hu, M., & Dahlgren, R. A. (2014). A dynamic watershed model for determining the effects of transient storage on nitrogen export to rivers. *Water Resources Research*, *50*(10), 7714–7730. doi: 10.1002/2014WR015852

Chen, D., Shen, H., Hu, M., Wang, J., Zhang, Y., & Dahlgren, R. A. (2018). Chapter Five - Legacy Nutrient Dynamics at the Watershed Scale: Principles, Modeling, and Implications. In D. L. Sparks (Ed.), *Advances in agronomy* (Vol. 149, pp. 237–313). Academic Press. doi: 10.1016/bs.agron.2018.01.005

Choi, H. T., & Beven, K. (2007). Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in an application of TOPMODEL within the GLUE framework. *Journal of Hydrology*, *332*(3-4), 316–336. doi: 10.1016/j.jhydrol.2006.07.012

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Rasmussen, R. M. (2015a). A unified approach for process-based hydrologic

modeling: 1. modeling concept. *Water Resources Research*, *51*(4), 2498–2514. doi: 10.1002/2015WR017198

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... Marks, D. G. (2015b). A unified approach for process-based hydrologic modeling: 2. model implementation and case studies. *Water Resources Research*, *51*(4), 2515–2542. doi: 10.1002/2015WR017200

Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., ... Wasson, M. F. (1999). Global patterns of terrestrial biological nitrogen (n2) fixation in natural ecosystems. *Global Biogeochemical Cycles*, *13*(2), 623–645. doi: 10.1029/1999GB900014

Damania, R., Desbureaux, S., Rodella, A.-S., Russ, J., & Zaveri, E. (2019). *Quality unknown: The invisible water crisis.* Washington, DC, USA: The World Bank. doi: 10.1596/978-1-4648-1459-4

Dehaspe, J., Sarrazin, F., Kumar, R., Fleckenstein, J. H., & Musolff, A. (2021). Bending of the concentration discharge relationship can inform about in-stream nitrate removal. *Hydrology and Earth System Sciences*, *25*(12), 6437–6463. doi: 10.5194/hess-25-6437-2021

Digizeitschriften. (2021). *Statistisches jahrbuch.* Retrieved from `http://www.digizeitschriften.de/dms/toc/?PID=PPN635628112` (last access: 07 November 2021)

Dupas, R., Ehrhardt, S., Musolff, A., Fovet, O., & Durand, P. (2020). Long-term nitrogen retention and transit time distribution in agricultural catchments in western France. *Environmental Research Letters*, *15*(11), 115011. doi: 10.1088/1748-9326/abbe47

DüV. (2007). *Düngeverordnung vom 27. Februar 2007 (BGBl. I S. 221).* Retrieved from `https://www.umwelt-online.de/recht/lebensmt/dvo_ges.htm` (last access: 28 January 2022)

DüV. (2017). *Düngeverordnung vom 26. Mai 2017 (BGBl. I S. 1305).* Retrieved from `http://www.gesetze-im-internet.de/d_v_2017/D%C3%BCV.pdf` (last access: 28 January 2022)

EEA. (2015). *Waterbase-UWWTD version 1.* European Environmental Agency (EEA) [data set]. Retrieved from `https://www.eea.europa.eu/ds_resolveuid/8396aa5079544dab9d8b8966237a9f3b`

EEA. (2018). *European waters: Assessment of status and pressures 2018* (No. European

Environment Agency (EEA) report 7/2018). Luxembourg, Luxembourg: Publications Office of the European Union. doi: 10.2800/303664

EEA. (2019a). *Corine Land Cover Version 20, Release date: 14-06-2019.* European Environment Agency (EEA) under the framework of the Copernicus Land Monitoring Service [data set]. Retrieved from `https://land.copernicus.eu/pan-european/corine-land-cover` (last access: 07 November 2021)

EEA. (2019b). *Nutrient enrichment and eutrophication in Europe's seas: Moving towards a healthy marine environment* (No. European Environment Agency (EEA) report 14/2019). Luxembourg, Luxembourg: Publications Office of the European Union. doi: 10.2800/092643

EEA. (2021). *Waterbase - Water Quality ICM.* European Environmental Agency (EEA) [data set]. Retrieved from `https://www.eea.europa.eu/ds_resolveuid/91fb02dc13cd485c81e8bbd048b5af8d` (last access: 28 January 2021)

Ehrhardt, S., Ebeling, P., Dupas, R., Kumar, R., Fleckenstein, J. H., & Musolff, A. (2021). Nitrate transport and retention in Western European catchments are shaped by hydroclimate and subsurface properties. *Water Resources Research*, *57*(10), e2020WR029469. doi: 10.1029/2020WR029469

Erisman, J. W., van Grinsven, H., Grizzetti, B., Bouraoui, F., Powlson, D., Sutton, M. A., ... Reis, S. (2011). The European nitrogen problem in a global perspective. In M. A. Sutton et al. (Eds.), *The european nitrogen assessment: Sources, effects and policy perspectives* (p. 9—31). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511976988.005

Eschenbach, W., Budziak, D., Elbracht, J., Höper, H., Krienen, L., Kunkel, R., ... Wendland, F. (2018). Möglichkeiten und Grenzen der Validierung flächenhaft modellierter Nitrateinträge ins Grundwasser mit der N2/Ar-Methode. *Grundwasser*, *23*, 125-–139. doi: 10.1007/s00767-018-0391-6

European Commission. (2020). *Tenth report on the implementation status and programmes for implementation (as required by Article 17 of Council Directive 91/271/EEC, concerning urban waste water treatment)* (No. COM/2020/492 final). Brussels, Belgium. Retrieved from `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0492` (last access: 07 November 2021)

Eurostat. (2016). *Population connected to wastewater collection and treatment systems by nuts 2 regions (env_wwcon_r2) - 11.04.2016 update.* European Commission [data set].

Retrieved from https://db.nomics.world/Eurostat/env_wwcon_r2 (last access: 07 November 2021)

Eurostat. (2021). *Population connected to wastewater treatment plants (env_ww_con) - 08-02-2021 update.* European Commission [data set]. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?lang=en&dataset=env_ww_con (last access: 07 November 2021)

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

FAO. (1951). *The state of food and agriculture 1951 - Review and overlook.* Rome, Italy: Food and Agriculture Organization of the United Nations (FAO). Retrieved from https://www.fao.org/publications/sofa/en/ (last access: 07 November 2021)

FAO. (2021a). *Food balances (old methodology and population) - 12.12.2017 update.* Rome, Italy: Food and Agriculture Organization of the United Nations (FAO) [data set]. Retrieved from http://www.fao.org/faostat/en/#data/FBSH (last access: 07 November 2021)

FAO. (2021b). *New food balances - 14.04.2021 update.* Rome, Italy: Food and Agriculture Organization of the United Nations (FAO) [data set]. Retrieved from http://www.fao.org/faostat/en/#data/FBS (last access: 07 November 2021)

FAO and SIK. (2011). *Global food losses and food waste - Extent, causes and prevention, Study conducted for the International Congress SAVE FOOD! at Interpack2011 Düsseldorf, Germany.* Rome, Italy: Food and Agriculture Organization of the United Nations (FAO).

Farzadfar, S., Knight, J. D., & Congreves, K. A. (2021). Soil organic nitrogen: an overlooked but potentially significant contribution to crop nutrition. *Plant and Soil*, *462*, 7—23. doi: 10.1007/s11104-021-04860-w

FGG Weser. (2020). *EG-Wasserrahmenrichtlinie, Die wichtigen Fragen der Gewässerbewirtschaftung in der Flussgebietseinheit Weser.* Hildesheim, Germany. Retrieved from https://www.fgg-weser.de/oeffentlichkeitsbeteiligung/veroeffentlichungen/eg-wrrl (last access: 01 March 2021)

FGG Weser. (2021). *Weser-Datenbank.* Hildesheim, Germany: Flussgebietsgemeinschaft (FGG) Weser [data set]. Retrieved from https://datenbank.fgg-weser.de/

`weserdatenbank` (last access: 07 November 2021)

Ford, W. I., Fox, J. F., & Pollock, E. (2017). Reducing equifinality using isotopes in a process-based stream nitrogen model highlights the flux of algal nitrogen from agricultural streams. *Water Resources Research*, *53*(8), 6539–6561. doi: 10.1002/2017WR020607

Franko, U., Oelschlägel, B., & Schenk, S. (1995). Simulation of temperature-, water- and nitrogen dynamics using the model CANDY. *Ecological Modelling*, *81*(1), 213–222. doi: 10.1016/0304-3800(94)00172-E

Global Runoff Data Centre. (2021). *Runoff data.* Koblenz, Germany: Bundesanstalt für Gewässerkunde [data set]. Retrieved from `http://www.bafg.de/GRDC` (last access: 10 February 2021)

Grimvall, A., Stålnacke, P., & Tonderski, A. (2000). Time scales of nutrient losses from land to sea — A European perspective. *Ecological Engineering*, *14*(4), 363–371. doi: 10.1016/S0925-8574(99)00061-0

Grizzetti, B., Bouraoui, F., & De Marsily, G. (2008). Assessing nitrogen pressures on european surface water. *Global Biogeochemical Cycles*, *22*(4), GB4023. doi: 10.1029/2007GB003085

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1), 80–91. doi: 10.1016/j.jhydrol.2009.08.003

Haag, D., & Kaupenjohann, M. (2001). Landscape fate of nitrate fluxes and emissions in Central Europe: A critical review of concepts, data, and models for transport and retention. *Agriculture, Ecosystems & Environment*, *86*(1), 1–21. doi: 10.1016/S0167-8809(00)00266-8

Hanel, M., Rakovec, O., Markonis, Y., Máca, P., Samaniego, L., Kyselý, J., & Kumar, R. (2018). Revisiting the recent european droughts from a long-term perspective. *Scientific Reports*, *8*, 9499. doi: 10.1038/s41598-018-27464-4

Hansen, S., Jensen, H. E., Nielsen, N. E., & Svendsen, H. (1991). Simulation of nitrogen dynamics and biomass production in winter wheat using the Danish simulation model DAISY. *Fertilizer Research*, *27*, 245-–259. doi: 10.1007/BF01051131

Hartmann, A., Barberá, J. A., & Andreo, B. (2017). On the value of water quality data and informative flow states in karst modelling. *Hydrology and Earth System Sciences*,

21(12), 5971–5985. doi: 10.5194/hess-21-5971-2017

Häußermann, U., Klement, L., Breuer, L., Ullrich, A., Wechsung, G., & Bach, M. (2020). Nitrogen soil surface budgets for districts in Germany 1995 to 2017. *Environmental Sciences Europe*, *32*(4), 109. doi: 10.1186/s12302-020-00382-x

Heidecke, C., Hirt, U., Kreins, P., Kuhr, P., Kunkel, R., Mahnkopf, J., ... Wendland, F. T. N. R. . (2015). *Endbericht zum Forschungsprojekt "Entwicklung eines Instrumentes für ein flussgebietsweites Nährstoffmanagement in der Flussgebietseinheit Weser" - AGRUM+ Weser.* Braunschweig, Germany: Johann Heinrich von Thünen-Institut. Retrieved from `https://www.thuenen.de/media/publikationen/thuenen-report/Thuenen-Report_21.pdf` (last access: 07 November 2021)

Hirsch, R. M., Archfield, S. A., & De Cicco, L. A. (2015). A bootstrap method for estimating uncertainty of water quality trends. *Environmental Modelling & Software*, *73*, 148-166. doi: 10.1016/j.envsoft.2015.07.017

Hirsch, R. M., Moyer, D. L., & Archfield, S. A. (2010). Weighted Regressions on Time, Discharge, and Season (WRTDS), with an application to Chesapeake Bay River inputs. *Journal of the American Water Resources Association*, *46*(5), 857–880. doi: 10.1111/j.1752-1688.2010.00482.x

Hirt, U., Kreins, P., Kuhn, U., Mahnkopf, J., Venohr, M., & Wendland, F. (2012). Management options to reduce future nitrogen emissions into rivers: A case study of the Weser river basin, Germany. *Agricultural Water Management*, *115*, 118–131. doi: 10.1016/j.agwat.2012.08.005

Holland, E. A., Dentener, F. J., Braswell, B. H., & Sulzman, J. M. (1999). Contemporary and pre-industrial global reactive nitrogen budgets. *Biogeochemistry*, *46*, 7–43. doi: 10.1007/BF01007572

Hong, B., Swaney, D. P., Mörth, C.-M., Smedberg, E., Eriksson Hägg, H., Humborg, C., ... Bouraoui, F. (2012). Evaluating regional variation of net anthropogenic nitrogen and phosphorus inputs (NANI/NAPI), major drivers, nutrient retention pattern and management implications in the multinational areas of Baltic Sea basin. *Ecological Modelling*, *227*, 117–135. doi: 10.1016/j.ecolmodel.2011.12.002

Howarth, R. W., Billen, G., Swaney, D., Townsend, A., Jaworski, N., Lajtha, K., ... Zhao-Liang, Z. (1996). Regional nitrogen budgets and riverine N & P fluxes for the drainages to the North Atlantic Ocean: Natural and human influences. *Biogeochemistry*, *35*, 75–139. doi: 10.1007/BF02179825

Husic, A., Fox, J., Adams, E., Ford, W., Agouridis, C., Currens, J., & Backus, J. (2019). Nitrate Pathways, Processes, and Timing in an Agricultural Karst System: Development and Application of a Numerical Model. *Water Resources Research*, *55*(3), 2079–2103. doi: 10.1029/2018WR023703

Ilampooranan, I., Van Meter, K. J., & Basu, N. B. (2019). A Race Against Time: Modeling Time Lags in Watershed Response. *Water Resources Research*, *55*(5), 3941–3959. doi: 10.1029/2018WR023815

Jenkinson, D. S. (1991). The Rothamsted long-term experiments: Are they still of use? *Agronomy Journal*, *83*(1), 2–10. doi: 10.2134/agronj1991.00021962008300010008x

Kaandorp, V. P., Broers, H. P., van der Velde, Y., Rozemeijer, J., & de Louw, P. G. B. (2021). Time lags of nitrate, chloride, and tritium in streams assessed by dynamic groundwater flow tracking in a lowland landscape. *Hydrology and Earth System Sciences*, *25*(6), 3691–3711. doi: 10.5194/hess-25-3691-2021

Kaplan, J. O., & Krumhardt, K. M. (2018). *The KK09 Anthropogenic Land Cover Change Scenarios for Europe and neighboring countries.* PANGEA [data set]. doi: 10.1594/PANGAEA.893758

Klein Goldewijk, K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene – HYDE 3.2. *Earth System Science Data*, *9*(2), 927–953. doi: 10.5194/essd-9-927-2017

Klein Goldewijk, K., Beusen, A., van Drecht, G., & de Vos, M. (2011). The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. *Global Ecology and Biogeography*, *20*(1), 73–86. doi: 10.1111/j.1466-8238.2010.00587.x

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. doi: 10.5194/hess-23-4323-2019

Knoll, L., Breuer, L., & Bach, M. (2020). Nation-wide estimation of groundwater redox conditions and nitrate concentrations through machine learning. *Environmental Research Letters*, *15*(6), 064004. doi: 10.1088/1748-9326/ab7d5c

Koeniger, P., Schwientek, M., Uhlenbrook, S., Leibundgut, C., & Krause, W. J. (2008). Tritium balance in macro-scale river basins analysed through distributed hydrological modelling. *Hydrological Processes*, *22*(5), 567–576. doi: 10.1002/hyp.6634

Kolmogorov, A. (1933). Sulla determinazione empirica di una legge di distribuzione. *Gior-*

nale dell'Istituto Italiano degli Attuari, *4*, 83–61.

Kreins, P., Behrendt, H., Gömann, H., Heidecke, C., Hirt, U., Kunkel, R., ... Wendland, F. (2010). *Analyse von Agrar- und Umweltmaßnahmen im Bereich des landwirtschaftlichen Gewässerschutzes vor dem Hintergrund der EG-Wasserrahmenrichtlinie in der Flussgebietseinheit Weser (special issue 336).* Braunschweig, Germany: Johann Heinrich von Thünen-Institut. Retrieved from `https://www.thuenen.de/media/publikationen/landbauforschung-sonderhefte/lbf_sh336.pdf` (last access: 07 November 2021)

Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resources Research*, *49*(1), 360–379. doi: 10.1029/2012WR012195

Lamarque, J.-F., Emmons, L. K., Hess, P. G., Kinnison, D. E., Tilmes, S., Vitt, F., ... Tyndall, G. K. (2012). CAM-chem: description and evaluation of interactive atmospheric chemistry in the Community Earth System Model. *Geoscientific Model Development*, *5*(2), 369–411. doi: 10.5194/gmd-5-369-2012

Lee, M., Shevliakova, E., Malyshev, S., Milly, P. C. D., & Jaffé, P. R. (2016). Climate variability and extremes, interacting with nitrogen storage, amplify eutrophication risk. *Geophysical Research Letters*, *43*(14), 7520–7528. doi: 10.1002/2016GL069254

Lewis, D. B., & Kaye, J. P. (2012). Inorganic nitrogen immobilization in live and sterile soil of old-growth conifer and hardwood forests: implications for ecosystem nitrogen retention. *Biogeochemistry*, *111*(1), 169-–186. doi: 10.1007/s10533-011-9627-6

Liu, J., Van Meter, K. J., McLeod, M. M., & Basu, N. B. (2021). Checkered landscapes: hydrologic and biogeochemical nitrogen legacies along the river continuum. *Environmental Research Letters*, *16*(11), 115006. doi: 10.1088/1748-9326/ac243c

Liu, X., Hu, G., He, H., Liang, C., Zhang, W., Bai, Z., ... Zhang, X. (2016). Linking microbial immobilization of fertilizer nitrogen to in situ turnover of soil microbial residues in an agro-ecosystem. *Agriculture, Ecosystems & Environment*, *229*, 40–47. doi: 10.1016/j.agee.2016.05.019

Lu, C., & Tian, H. (2017). Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth System Science Data*, *9*(1), 181-192. doi: 10.5194/essd-9-181-2017

Macdonald, A. J., Powlson, D. S., Poulton, P. R., & Jenkinson, D. S. (1989). Unused fertiliser nitrogen in arable soils - Its contribution to nitrate leaching. *Journal of the*

Science of Food and Agriculture, *46*(4), 407–419. doi: 10.1002/jsfa.2740460404

MacDonald, G., Levison, J., & Parker, B. (2017). On Methods for In-Well Nitrate Monitoring Using Optical Sensors. *Groundwater Monitoring & Remediation*, *37*(4), 60–70. doi: 10.1111/gwmr.12248

Mariotti, F., Tomé, D., & Mirand, P. P. (2008). Converting nitrogen into protein—beyond 6.25 and jones' factors. *Critical Reviews in Food Science and Nutrition*, *48*(2), 177–184. doi: 10.1080/10408390701279749

Martinez, G. F., & Gupta, H. V. (2010). Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous united states. *Water Resources Research*, *46*(8), W08507. doi: 10.1029/2009WR008294

Morée, A. L., Beusen, A. H. W., Bouwman, A. F., & Willems, W. J. (2013). Exploring global nitrogen and phosphorus flows in urban wastes during the twentieth century. *Global Biogeochemical Cycles*, *27*(3), 836–846. doi: 10.1002/gbc.20072

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, *58*(6), 1763–1785. doi: 10.13031/trans.58.10715

Moriasi, D. N., Zeckoski, R. W., Arnold, J. G., Baffaut, C., Malone, R. W., Daggupati, P., ... Douglas-Mankin, K. R. (2015). Hydrologic and Water Quality Models: Key Calibration and Validation Topics. *Transactions of the ASABE*, *58*(6), 1609–1618. doi: 10.13031/trans.58.11075

Morier, I., Schleppi, P., Siegwolf, R., Knicker, H., & Guenat, C. (2008). 15N immobilization in forest soil: a sterilization experiment coupled with 15CPMAS NMR spectroscopy. *European Journal of Soil Science*, *59*(3), 467—475. doi: 10.1111/j.1365-2389.2007.00998.x

Mulholland, P. J., Helton, A. M., Poole, G. C., Hall, R. O., Hamilton, S. K., Peterson, B. J., ... Thomas, S. M. (2008). Stream denitrification across biomes and its response to anthropogenic nitrate loading. *Nature*, *452*(7184), 202–205. doi: 10.1038/nature06686

Myhre, D., G., Shindell, F.-M. B., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., ... Zhang, H. (2013). Chapter 8. Anthropogenic and Natural Radiative Forcing. In T. F. Stocker et al. (Eds.), *Climate change 2013: The physical science basis. Working Group I contribution to the fifth assessment report of the Intergovernmental Panel on Climate Change* (pp. 659–740). New York, USA: Cambridge University Press.

Näsholm, T., Kielland, K., & Ganeteg, U. (2009). Uptake of organic nitrogen by plants. *New Phytologist*, *182*(1), 31–48. doi: 10.1111/j.1469-8137.2008.02751.x

Nguyen, T. V., Kumar, R., Lutz, S. R., Musolff, A., Yang, J., & Fleckenstein, J. H. (2021). Modeling Nitrate Export From a Mesoscale Catchment Using StorAge Selection Functions. *Water Resources Research*, *57*(2), e2020WR028490. doi: 10.1029/2020WR028490

NLWKN. (2019). *Grundwasserbericht Niedersachsen: Kurzbericht 2019: Grundwasserstand sowie Güteparameter Nitrat und Arzneimittel im Grundwasser (Band 38).* Hildesheim/Norden, Germany: Niedersächsischer Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN). Retrieved from `https://www.nlwkn.niedersachsen.de/startseite/wasserwirtschaft/grundwasser/veroffentlichungen/publikationsreihe_grundwasser/veroeffentlichungen-zum-thema-grundwassertrinkwasser-zum-downloaden-198537.html` (last access: 07 October 2021)

NLWKN. (2022). *Landesweite Datenbank für wasserwirtschaftliche Daten.* Norden, Germany: Niedersächsische Landesbetrieb für Wasserwirtschaft, Küsten- und Naturschutz (NLWKN) [data set]. Retrieved from `http://www.wasserdaten.niedersachsen.de/cadenza/` (last access: 28 January 2022)

OGewV. (2016). *Oberflächengewässerverordnung vom 20. Juni 2016 (BGBl. I S. 1373), die zuletzt durch Artikel 2 Absatz 4 des Gesetzes vom 9. Dezember 2020 (BGBl. I S. 2873) geändert worden ist.* Retrieved from `https://www.gesetze-im-internet.de/ogewv_2016/index.html` (last access: 07 November 2021)

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E. C., van Emmerik, T., Uijlenhoet, R., ... Woods, R. (2017). Scaling, similarity, and the fourth paradigm for hydrology. *Hydrology and Earth System Sciences*, *21*(7), 3701–3713. doi: 10.5194/hess-21-3701-2017

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214-232. doi: 10.1016/j.envsoft.2016.02.008

Pianosi, F., Sarrazin, F., & Wagener, T. (2015). A Matlab toolbox for Global Sensitivity Analysis. *Environmental Modelling & Software*, *70*, 80–85. doi: 10.1016/j.envsoft.2015.04.009

Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, *67*, 1–11. doi: 10.1016/j.envsoft.2015.01.004

Pianosi, F., & Wagener, T. (2018). Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling & Software*, *108*, 197–207. doi: 10.1016/j.envsoft.2018.07.019

Poisvert, C., Curie, F., & Moatar, F. (2017). Annual agricultural N surplus in France over a 70-year period. *Nutrient Cycling in Agroecosystems*, *107*(9), 63–78. doi: 10.1007/s10705-016-9814-x

Portmann, R. W., Daniel, J. S., & Ravishankara, A. R. (2012). Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *367*(1593), 1256—1264. doi: 10.1098/rstb.2011.0377

Puckett, L. J., Tesoriero, A. J., & Dubrovsky, N. M. (2011). Nitrogen contamination of surficial aquifers — A growing legacy. *Environmental Science & Technology*, *45*(3), 839–844. doi: 10.1021/es1038358

Queloz, P., Carraro, L., Benettin, P., Botter, G., Rinaldo, A., & Bertuzzo, E. (2015). Transport of fluorobenzoate tracers in a vegetated hydrologic control volume: 2. Theoretical inferences and modeling. *Water Resources Research*, *51*(4), 2793–2806. doi: 10.1002/2014WR016508

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., ... Samaniego, L. (2016). Multiscale and multivariate evaluation of water fluxes and states over European river basins. *Journal of Hydrometeorology*, *17*(1), 287–307. doi: 10.1175/JHM-D-15-0054.1

Rankinen, K., Karvonen, T., & Butterfield, D. (2006). An application of the GLUE methodology for estimating the parameters of the INCA-N model. *Science of The Total Environment*, *365*(1), 123–139. doi: 10.1016/j.scitotenv.2006.02.034

Rivett, M. O., Buss, S. R., Morgan, P., Smith, J. W., & Bemment, C. D. (2008). Nitrate attenuation in groundwater: A review of biogeochemical controlling processes. *Water Research*, *42*(16), 4215–4232. doi: 10.1016/j.watres.2008.07.020

Robertson, G. P., & Groffman, P. M. (2015). Chapter 14 - nitrogen transformations. In E. A. Paul (Ed.), *Soil microbiology, ecology and biochemistry* (fourth ed., pp. 421–446). Boston, USA: Academic Press. doi: 10.1016/B978-0-12-415955-6.00014-1

Rudolph, D. L., Barry, D. A. J., & Goss, M. J. (1998). Contamination in Ontario farmstead

domestic wells and its association with agriculture: 2. Results from multilevel monitoring well installations. *Journal of Contaminant Hydrology*, *32*(3), 295–311. doi: 10.1016/S0169-7722(98)00053-9

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5), W05523. doi: 10.1029/2008WR007327

Samaniego, L., Kumar, R., Thober, S., Rakovec, O., Zink, M., Wanders, N., ... Attinger, S. (2017). Toward seamless hydrologic predictions across spatial scales. *Hydrology and Earth System Sciences*, *21*(9), 4323–4346. doi: 10.5194/hess-21-4323-2017

Sarrazin, F., Hartmann, A., Pianosi, F., Rosolem, R., & Wagener, T. (2018). V2Karst V1.1: a parsimonious large-scale integrated vegetation–recharge model to simulate the impact of climate and land cover change in karst regions. *Geoscientific Model Development*, *11*(12), 4933–4964. doi: 10.5194/gmd-11-4933-2018

Sarrazin, F., Pianosi, F., & Wagener, T. (2016). Global sensitivity analysis of environmental models: Convergence and validation. *Environmental Modelling & Software*, *79*, 135–152. doi: 10.1016/j.envsoft.2016.02.005

Scavia, D., & Bricker, S. (2006). Coastal eutrophication assessment in the United States. *Biogeochemistry*, *79*, 187—208. doi: 10.1007/s10533-006-9011-0

Sebilo, M., Mayer, B., Nicolardot, B., Pinay, G., & Mariotti, A. (2013). Long-term fate of nitrate fertilizer in agricultural soils. *Proceedings of the National Academy of Sciences*, *110*(45), 18185–18189. doi: 10.1073/pnas.1305372110

Seeger, H. (1999). The history of German waste water treatment. *European Water Management*, *2*(5). Retrieved from `http://www.sewerhistory.orgwww.sewerhistory.org/articles/whregion/germany/Seeger.pdf` (last access: 07 November 2021)

Six, J., Conant, R. T., Paul, E. A., & Paustian, K. (2002). Stabilization mechanisms of soil organic matter: Implications for C-saturation of soils. *Plant and Soil*, *241*, 155—176. doi: 10.1023/A:1016125726789

Smil, V. (1999). Nitrogen in crop production: An account of global flows. *Global Biogeochemical Cycles*, *13*(2), 647–662. doi: 10.1029/1999GB900015

Smirnov, N. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bulletin Mathématique de l'Université Moscou*, *2*, 3–14.

Spoelstra, J., Schiff, S. L., Elgood, R. J., Semkin, R. G., & Jeffries, D. S. (2001). Tracing

the Sources of Exported Nitrate in the Turkey Lakes Watershed Using 15N/14N and 18O/16O isotopic ratios. *Ecosystems*, *4*, 536—544. doi: 10.1007/s10021-001-0027-y

Statistisches Bundesamt. (2021). *Online data portal Genesis, Tables 41120-01-02-4-B and 41141-01-01-4-B.* Wiesbaden, Germany: Gemeinsames Neues Statistisches Informations-System (GENESIS) [data set]. Retrieved from `https://www.regionalstatistik.de/genesis/online` (last access: 07 November 2021)

Stuart, M., Chilton, P., Kinniburgh, D., & Cooper, D. (2007). Screening for long-term trends in groundwater nitrate monitoring data. *Quarterly Journal of Engineering Geology and Hydrogeology*, *40*(4), 361-–376. doi: 10.1144/1470-9236/07-040

Sültenfuß, J., Roether, W., & Rhein, M. (2009). The bremen mass spectrometric facility for the measurement of helium isotopes, neon, and tritium in water. *Isotopes in Environmental and Health Studies*, *45*(2), 83–95. doi: 10.1080/10256010902871929

Sundermann, G., Wägner, N., Cullmann, A., von Hirschhausen, C., & Kemfert, C. (2020). Nitrate pollution of groundwater long exceeding trigger value: fertilization practices require more transparency and oversight. *DIW weekly report*, *8+9/2020*, 61–72. doi: 10.18723/diw_dwr:2020-8-1

Svirejeva-Hopkins, A., Reis, S., Magid, J., Nardoto, G. B., Barles, S., Bouwman, A. F., ... Sutton, M. A. (2011). Nitrogen flows and fate in urban landscapes. In M. A. Sutton et al. (Eds.), *The European Nitrogen Assessment: Sources, Effects and Policy Perspectives* (p. 249—270). Cambridge, UK: Cambridge University Press. doi: 10.1017/CBO9780511976988.015

Tian, H., Yang, J., Lu, C., Xu, R., Canadell, J. G., Jackson, R. B., ... Zhu, Q. (2018). The global N2O model intercomparison project. *Bulletin of the American Meteorological Society*, *99*(6), 1231–1251. doi: 10.1175/BAMS-D-17-0212.1

Tilmes, S., Lamarque, J.-F., Emmons, L. K., Kinnison, D. E., Marsh, D., Garcia, R. R., ... Blake, N. (2016). Representation of the Community Earth System Model (CESM1) CAM4-chem within the Chemistry-Climate Model Initiative (CCMI). *Geoscientific Model Development*, *9*(5), 1853–1890. doi: 10.5194/gmd-9-1853-2016

United Nations. (2017). *The United Nations World Water Development Report 2017, Wastewater: the untapped resource.* Paris, France: UNESCO. Retrieved from `http://www.unesco.org/new/en/natural-sciences/environment/water/wwap/wwdr/2017-wastewater-the-untapped-resource/` (last access: 07 November 2021)

Van Drecht, G., Bouwman, A. F., Harrison, J., & Knoop, J. M. (2009). Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050. *Global Biogeochemical Cycles*, *23*(4). doi: 10.1029/2009GB003458

Van Meter, K. J., & Basu, N. B. (2015). Catchment legacies and time lags: a parsimonious watershed model to predict the effects of legacy storage on nitrogen export. *PLoS ONE*, *10*, e0125971. doi: 10.1371/journal.pone.0125971

Van Meter, K. J., & Basu, N. B. (2017). Time lags in watershed-scale nutrient transport: an exploration of dominant controls. *Environmental Research Letters*, *12*(8), 084017. doi: 10.1088/1748-9326/aa7bf4

Van Meter, K. J., Basu, N. B., & Van Cappellen, P. (2017). Two centuries of nitrogen dynamics: Legacy sources and sinks in the Mississippi and Susquehanna River Basins. *Global Biogeochemical Cycles*, *31*(1), 2–23. doi: 10.1002/2016GB005498

Van Meter, K. J., Basu, N. B., Veenstra, J. J., & Burras, C. L. (2016). The nitrogen legacy: Emerging evidence of nitrogen accumulation in anthropogenic landscapes. *Environmental Research Letters*, *11*(3), 035014. doi: 10.1088/1748-9326/11/3/035014

Van Meter, K. J., Van Cappellen, P., & Basu, N. B. (2018). Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico. *Science*, *360*(6387), 427–430. doi: 10.1126/science.aar4462

Van Puijenbroek, P. J. T. M., Beusen, A. H. W., & Bouwman, A. F. (2019). Global nitrogen and phosphorus in urban waste water based on the shared socio-economic pathways. *Journal of Environmental Management*, *231*, 446–456. doi: 10.1016/j.jenvman.2018.10.048

Vero, S., Basu, N., Van Meter, K., Richards, K. G., Mellander, P.-E., Healy, M. G., & Fenton, O. (2018). Review: the environmental status and implications of the nitrate time lag in Europe and North America. *Hydrogeology Journal*, *26*, 7—22. doi: 10.1007/s10040-017-1650-9

Wade, A., Jackson, B., & Butterfield, D. (2008). Over-parameterised, uncertain 'mathematical marionettes' — How can we best use catchment water quality models? An example of an 80-year catchment-scale nutrient balance. *Science of The Total Environment*, *400*(1), 52–74. doi: 10.1016/j.scitotenv.2008.04.030

Wang, L., Stuart, M. E., Bloomfield, J. P., Butcher, A. S., Gooddy, D. C., McKenzie, A. A., ... Williams, A. T. (2012). Prediction of the arrival of peak nitrate concentrations at the water table at the regional scale in Great Britain. *Hydrological Processes*, *26*(2),

226–239. doi: 10.1002/hyp.8164

Wang, X., Feng, A., Wang, Q., Wu, C., Liu, Z., Ma, Z., & Wei, X. (2014). Spatial variability of the nutrient balance and related NPSP risk analysis for agro-ecosystems in China in 2010. *Agriculture, Ecosystems & Environment*, *193*, 42–52. doi: 10.1016/j.agee.2014.04.027

Wendland, F., Behrendt, H., Gömann, H., Hirt, U., Kreins, P., Kuhn, U., ... Tetzlaff, B. (2009). Determination of nitrogen reduction levels necessary to reach groundwater quality targets in large river basins: the Weser basin case study, Germany. *Nutrient Cycling in Agroecosystems*, *85*, 63–78. doi: 10.1007/s10705-009-9248-9

WHO. (2016). *Nitrate and nitrite in drinking-water - Background document for development of WHO Guidelines for Drinking-water Quality* (No. WHO/FWC/WSH/16.52). Geneva , Switzerland: World Health Organization (WHO) Press. Retrieved from `https://www.who.int/water_sanitation_health/dwq/chemicals/nitrate-nitrite-background-jan17.pdf` (last access: 07 November 2021)

Worrall, F., Howden, N., & Burt, T. (2015). Evidence for nitrogen accumulation: The total nitrogen budget of the terrestrial biosphere of a lowland agricultural catchment. *Biogeochemistry*, *123*, 411—428. doi: 10.1007/s10533-015-0074-7

Wriedt, G., & Rode, M. (2006). Modelling nitrate transport and turnover in a lowland catchment system. *Journal of Hydrology*, *328*(1), 157–176. doi: 10.1016/j.jhydrol.2005.12.017

Yang, S., Büttner, O., Kumar, R., Jäger, C., Jawitz, J. W., Rao, P., & Borchardt, D. (2019). Spatial patterns of water quality impairments from point source nutrient loads in Germany's largest national River Basin (Weser River). *Science of The Total Environment*, *697*, 134145. doi: 10.1016/j.scitotenv.2019.134145

Yang, X., Jomaa, S., Zink, M., Fleckenstein, J. H., Borchardt, D., & Rode, M. (2018). A new fully distributed model of nitrate transport and removal at catchment scale. *Water Resources Research*, *54*(8), 5856–5877. doi: 10.1029/2017WR022380

Yansheng, C., Fengliang, Z., Zhongyi, Z., Tongbin, Z., & Huayun, X. (2020). Biotic and abiotic nitrogen immobilization in soil incorporated with crop residue. *Soil and Tillage Research*, *202*, 104664. doi: 10.1016/j.still.2020.104664

Zhang, B., Tian, H., Lu, C., Dangal, S. R. S., Yang, J., & Pan, S. (2017). Global manure nitrogen production and application in cropland during 1860–2014: a 5 arcmin gridded

global dataset for Earth system modeling. *Earth System Science Data*, *9*(2), 667–678. doi: 10.5194/essd-9-667-2017

Zhang, X., Zou, T., Lassaletta, L., Mueller, N. D., Tubiello, F. N., Lisk, M. D., ... Davidson, E. A. (2021). Quantification of global and national nitrogen budgets for crop production. *Nature Food*, *2*, 529—540. doi: 10.1038/s43016-021-00318-5

Zink, M., Kumar, R., Cuntz, M., & Samaniego, L. (2017). A high-resolution dataset of water fluxes and states for germany accounting for parametric uncertainty. *Hydrology and Earth System Sciences*, *21*(3), 1769–1790. doi: 10.5194/hess-21-1769-2017

Supporting Information for "Characterizing Catchment Scale Nitrogen Legacies and Constraining their Uncertainties"

F. J. Sarrazin [1], R. Kumar[1], N. B. Basu[2,3,4], A. Musolff[5], M. Weber[1], K. Van Meter[6,7], S. Attinger[1,8]

[1]Department of Computational Hydrosystems, Helmholtz-Centre for Environmental Research - UFZ, Leipzig, Germany

[2]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada

[3]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada

[4]Water Institute, University of Waterloo, Waterloo, ON, Canada

[5]Department of Hydrogeology, Helmholtz-Centre for Environmental Research - UFZ, Leipzig, Germany

[6]Department of Geography, Pennsylvania State University, PA, USA

[7]Earth and Environmental Systems Institute, Pennsylvania State University, PA, USA

[8]Institute of Environmental Science and Geography, University of Potsdam, Potsdam, Germany

Contents of this file

Section S1. Derivation of the Equation for the Subsurface Compartment in ELEMeNT

In this section we provide details on the derivation of the equation for the subsurface compartment (Equation 1 in the main article):

$$J_{stream_{sub}}(t) = \int_0^{+\infty} J_{sub_s}(t-T)p(T,t-T)e^{-\lambda_{sub}T}dT \tag{S1}$$

The travel time distribution *p(T,t-T)* of Equation S1 refers to the concept of forward travel time, where the travel time T corresponds to the life expectancy of the water particles, and (t-T) represents their time of injection into the subsurface compartment. We can show that Equation S1 is equivalent to the widely used equation which links the solute (N) concentration in the outflow (here the concentration in the outflow of the subsurface compartment, denoted as $C_{stream_{sub}}$ $(mg\ L^{-1})$) to the solute (N) concentration in the storage (see e.g. Equation 9 in Benettin et al., 2015):

$$C_{stream_{sub}}(t) = \int_0^{+\infty} C_{sub}(T)p_b(T,t)dT \tag{S2}$$

In Equation S2, $p_b(T,t)$ (-) is the backward travel time distribution, where the travel time T represents the age of the water particles in the storage, and $C_{sub}(T)$ $(mg\ L^{-1})$ is the solute (N) concentration of the water particles in the storage with age T.

Equation S2 can be rewritten as a function of the concentration in the inflow at time t-T, here the catchment-scale concentration in the percolating water from the source zone to the subsurface denoted as $C_{sub_s}(t-T)$ $(mg\ L^{-1})$, and a decay/degradation function, here accounting for denitrification in the subsurface (Queloz et al., 2015):

$$C_{stream_{sub}}(t) = \int_0^{+\infty} C_{sub_s}(t-T)e^{-\lambda_{sub}T}p_b(T,t)dT \tag{S3}$$

Furthermore, forward and backward travel time distributions are linked by the following relationship (see e.g. Equation 16 in Benettin et al., 2015):

$$p_b(T,t) = \frac{Q_{sub_s}(t-T)}{Q_{stream_{sub}}(t)} p(T, t-T) \tag{S4}$$

where $Q_{sub_s}(t)$ $(mm\ yr^{-1})$ is the inflow from the source zone to the subsurface and $Q_{stream_{sub}}(t)$ $(mm\ yr^{-1})$ is the outflow from the subsurface to the stream.

Therefore, Equation S3 can be further modified as:

$$C_{stream_{sub}}(t) Q_{stream_{sub}}(t) = \int_0^{+\infty} C_{sub_s}(t-T) Q_{sub_s}(t-T) p(T, t-T) e^{-\lambda_{sub} T} dT \tag{S5}$$

Equation S5 is equivalent to Equation S1.

We note that in ELEMeNT, it is assumed that for any given time t the subsurface outflow, the subsurface inflow and the stream discharge at the catchment outlet Q_{out}(t) $(mm\ yr^{-1})$ are equal, i.e.:

$$Q_{stream_{sub}}(t) = Q_{sub_s}(t) = Q_{out}(t) \tag{S6}$$

Section S2. Derivation of the Mean Travel Time for the Subsurface Compartment in ELEMeNT

In this section we provide details on the derivation of Equations 2-3 in the main article:

$$p(T, t-T) = \frac{1}{\mu'(t)} e^{-\int_{t-T}^{t} \frac{1}{\mu'(x)} dx} \tag{S7}$$

$$\mu'(t) = \frac{\overline{Q_{out}}\mu_{sub}}{Q_{out}(t)} \tag{S8}$$

Under assumption of complete mixing in the subsurface storage (or of random sampling of the water particles from the storage), the forward travel time distribution used in Equation S1 can be expressed as Equation 41 in Botter, Bertuzzo, and Rinaldo (2010), considering that no evapotranspiration occurs in the subsurface storage:

$$p(T, t-T) = \frac{Q_{out}(t)}{V_{sub}} e^{-\int_{t-T}^{t} \frac{Q_{out}(x)}{V_{sub}} dx} \tag{S9}$$

where V_{sub} (mm) is the depth of the subsurface storage, which is constant in time in ELEMeNT, since it is assumed that the inflow is equal to the outflow as reported in Equation S6. Hence, from Equation S9, we can express the parameter of the travel time distribution $\mu'(t)$ (yr) as:

$$\frac{1}{\mu'(t)} = \frac{Q_{out}(t)}{V_{sub}} \tag{S10}$$

By averaging Equation S10 over time we obtain:

$$V_{sub} = \overline{Q_{out}}\mu_{sub} \tag{S11}$$

where μ_{sub} (yr) is the harmonic mean of $\mu'(t)$ and $\overline{Q_{out}}$ ($mm\ yr^{-1}$) is the arithmetic mean of Q_{out}. By combining Equation S11 and Equation S10, we obtain Equation S8.

Section S3. Numerical Implementation of the Equations for the Subsurface Compartment in ELEMeNT

The dynamic of the N mass stored in the subsurface compartment $M_{sub}(t)$ $(kg\ ha^{-1})$ is governed by the following differential equation:

$$\frac{dM_{sub}}{dt}(t) = J_{sub_s}(t) - \lambda_{sub} M_{sub}(t) - \frac{1}{\mu'(t)} M_{sub}(t) \tag{S12}$$

We note that Equation S1 (Equation 1 in the main article) is the integrated form of Equation S12. To simulate the ELEMeNT model, Equation S12 is solved numerically. The denitrification flux $J_{den_{sub}}(t_{i-1} \to t_i)$ $(kg\ ha^{-1}\ yr^{-1})$ and the N mass flux exported to the stream $J_{stream_{sub}}(t_{i-1} \to t_i)$ $(kg\ ha^{-1}\ yr^{-1})$ during the i-th simulation time step denoted as $]t_{i-1}, t_i]$, and the N mass stored in the subsurface compartment $M_{sub}(t_i)$ $(kg\ ha^{-1})$ at the end of the i-th time step, are computed as reported below:

$$M_{sub}(t_i) = \left(M_{sub}(t_{i-1}) - \frac{J_{sub_s}(t_{i-1} \to t_i)}{\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \to t_i)}} \right) e^{\left(\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \to t_i)}\right)(t_{i-1} - t_i)} + \frac{J_{sub_s}(t_{i-1} \to t_i)}{\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \to t_i)}} \tag{S13}$$

$$J_{den_{sub}}(t_{i-1} \to t_i) = \frac{\lambda_{sub}\mu'(t_{i-1} \to t_i)}{\lambda_{sub}\mu'(t_{i-1} \to t_i) + 1} \left(\frac{M_{sub}(t_i) - M_{sub}(t_{i-1})}{t_i - t_{i-1}} + J_{sub_s}(t_{i-1} \to t_i) \right) \tag{S14}$$

$$J_{stream_{sub}}(t_{i-1} \to t_i) = \frac{1}{\lambda_{sub}\mu'(t_{i-1} \to t_i) + 1} \left(\frac{M_{sub}(t_i) - M_{sub}(t_{i-1})}{t_i - t_{i-1}} + J_{sub_s}(t_{i-1} \to t_i) \right) \tag{S15}$$

In the following, we provide details on the derivation of Equations S13-S15. To solve Equation S12 numerically, we perform piecewise integration over each simulation time step $]t_{i-1}, t_i]$. For the integration, we assume that the rate of N leaching from the source zone to the subsurface, as well as the mean travel time are constant over each simulation time interval. We also consider that the function $M_{sub}(t)$ is continuous over the simulation period. The conditions for the integration over the i-th simulation time step can be

summarized as follows:

$$\begin{cases} J_{sub_s}(t) = J_{sub_s}(t_{i-1} \rightarrow t_i) \ \forall t \in]t_{i-1}, t_i] \\ Q_{out}(t) = Q_{out}(t_{i-1} \rightarrow t_i) \ \forall t \in]t_{i-1}, t_i] \\ \lim\limits_{t \rightarrow t_{i-1}} M_{sub}(t) = M_{sub}(t_{i-1}) \end{cases} \tag{S16}$$

Considering the conditions of Equation S16, we can solve Equation S12 and we can then write for a given t in $]t_{i-1}, t_i]$:

$$M_{sub}(t) = \left(M_{sub}(t_{i-1}) - \frac{J_{sub_s}(t_{i-1} \rightarrow t_i)}{\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \rightarrow t_i)}} \right) e^{\left(\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \rightarrow t_i)}\right)(t_{i-1} - t)} + \frac{J_{sub_s}(t_{i-1} \rightarrow t_i)}{\lambda_{sub} + \frac{1}{\mu'(t_{i-1} \rightarrow t_i)}} \tag{S17}$$

From Equation S17, we can derive Equation S13. The total mass flux leaving the subsurface compartment $J_{tot_{sub}}(t_{i-1} \rightarrow t_i)$ $(kg\ ha^{-1}\ yr^{-1})$, which is the sum of the denitrification flux and the N mass flux exported to the stream, can be computed numerically using the mass balance equation:

$$J_{tot_{sub}}(t_{i-1} \rightarrow t_i) = \frac{M_{sub}(t_i) - M_{sub}(t_{i-1})}{t_i - t_{i-1}} + J_{sub_s}(t_{i-1} \rightarrow t_i) \tag{S18}$$

Finally, the denitrification flux and the N mass flux exported to the stream are given as:

$$J_{den_{sub}}(t_{i-1} \rightarrow t_i) = \frac{\lambda_{sub}\mu'(t_{i-1} \rightarrow t_i)}{\lambda_{sub}\mu'(t_{i-1} \rightarrow t_i) + 1} J_{tot_{sub}}(t_{i-1} \rightarrow t_i) \tag{S19}$$

$$J_{stream_{sub}}(t_{i-1} \rightarrow t_i) = \frac{1}{\lambda_{sub}\mu'(t_{i-1} \rightarrow t_i) + 1} J_{tot_{sub}}(t_{i-1} \rightarrow t_i) \tag{S20}$$

Hence Equations S14 and S15.

Section S4. Construction of the Land Use Data

We construct the 1800–2015 trajectories of the catchment-scale fractions of the three land use categories required by ELEMeNT, namely cropland, agricultural permanent grassland and non-agricultural land (which includes forest, natural grassland, green urban areas, built-up areas and non-vegetated land).

From the Corine Land Cover dataset (CLC; EEA, 2019), we derive the fraction of forest (classes 311 to 313) and non-vegetated land (classes 331 to 523), since CLC is the reference land cover product for European countries. The CLC dataset provides maps for five years (1990, 2000, 2006, 2012 and 2018). In addition, we verify the consistency of the CLC forest fraction at the country scale (around 29%–30%) with the national forest inventories of 2002 and 2012 (around 32%; BMEL, 2014). For the period 1990–2015 we make use of the five CLC maps and fill the values for the years in-between using linear interpolation. For the period 1800–1990, which is not covered by the CLC dataset, and for which, to our knowledge, no inventory data are available, we consider no changes in forest and non-vegetated areas. This assumption is supported by the land cover reconstruction dataset of Kaplan and Krumhardt (2018), according to which the forest fraction in Germany is almost constant and around 34–35% for the period 1600–1850.

We do not extract agricultural areas from CLC, which are greatly overestimated as reported e.g. in Bach et al. (2006). Rather, we use the HYDE dataset (History Database of the Global Environment; Klein Goldewijk et al., 2011, 2017) to identify cropland, agricultural permanent grassland (called "grazing" in the HYDE dataset), as well as build-up areas. HYDE provides a consistent long-term time series of land use fractions covering the period 1800–2015. For the agricultural areas, we use the spatial distribution from

HYDE and we adjust the actual values to match census data. Census data is available at the county level for the period 1999–2016 from the Federal Statistical Office (Statistisches Bundesamt, 2021) and at the state level for the earlier period from the yearly statistical books (Digizeitschriften, 2021). We fill with linear interpolation the values in the HYDE dataset and the census data for the years for which values are not provided.

Finally, we attribute the remaining fraction of the land to a land use category that we call "other vegetated land", which includes in particular natural grassland, urban parks, and green areas in discontinuous urban fabric. No land use inventory allows us to distinguish further land use categories within this "other vegetated land" class. The trajectories of the land use fractions for the different subcatchments are presented in supplementary Figure S1. We observe that the uncertainty in the land use fractions, resulting from different scenarios provided by the HYDE dataset, is relatively small for the period 1850–2015 (Figure S1). Therefore we only consider the baseline scenario of HYDE to force the ELEMeNT model.

Section S5. Construction of the N Surplus for Agricultural Areas

We harmonize the two N surplus datasets of Häußermann et al. (2020) and Behrendt et al. (2003), similar to Ehrhardt et al. (2021), to construct the N surplus trajectories at county level for the period 1950–2015. During the overlapping period between the two datasets (1995–1998), we find that on average the state level N surplus values from Behrendt et al. (2000) underestimate the values derived from Häußermann et al. (2020) by 5-10%. We bias correct the N surplus values provided by Behrendt et al. (2000) with the values of Häußermann et al. (2020) for consistency. We then downscale the 1950–1995 bias-corrected state level values to county level assuming that, for all the counties within a given state, the N surplus followed the same temporal dynamics, while ensuring that the state totals are satisfied. The resulting trajectories of the N surplus for agricultural areas are shown in Figure S2.

Section S6. Construction of the N Point Sources

N point sources for the period 1950–2015 are constructed from the methodology proposed by Morée et al. (2013) (see Figure S4). N gross emissions from households are calculated from protein supply data (FAO, 1951, 2021a, 2021b), considering protein losses at the retail and household level (FAO and SIK, 2011) and N losses in humans via sweat, hair and blood (Morée et al., 2013). A portion of the household N gross emissions is collected by the sewer system, part of it being treated in waste water treatment plants (WWTPs) that can have different levels of efficiency (primary, secondary or tertiary treatment). The population connection to sewer and WWTPs data come from Seeger (1999) for the period before 1990 and Eurostat (2016, 2021) for the period after 1990. Following Morée et al. (2013), industrial N gross emissions are assumed to be equal to a (calibrated) fraction of the household N emissions. A fraction of the industrial N emissions is treated in WWTPs, while the other part ends up in stabilization pounds or is lost via volatilization. The total N point sources is computed as the sum of the N loading corresponding to the untreated fraction of the household N emissions collected by the sewer system, and the N loading coming from WWTPs.

The construction of the N point sources data requires the calibration of a number of coefficients that are described in Table S2. To account for uncertainty in these coefficients, we generate a ensemble of 100,000 combinations of values using latin hypercube sampling, uniform distribution and the ranges reported in Table S2, from which we obtain an ensemble of 100,000 realizations of the N point sources. We then select the 100 "best" realizations, i.e., the realizations for which the calculated N loading from WWTPs shows the smallest error with respect to the observation data of Büttner (2020). Further details

on the datasets used are reported in Table S1. A visual depiction of the N point sources with uncertainty is provided in Figure S5.

Section S7. N Pools Initial Conditions in ELEMeNT

Source Zone Initial States

The choice of the value of the initial N storage for the organic stores needs to be carefully examined, since the residence time of N in the organic pools can be large (Van Meter et al., 2017). In addition, the source zone compartment of ELEMeNT has a semi-distributed representation, as the N mass balance is evaluated over a number of units (100 in this study) that have no explicit spatial location and that have distinct land use trajectories. Therefore, the initial value of the organic N storages needs to be specified for the three land use types, i.e., cropland, agricultural permanent grassland and non-agricultural land. We further note that the ELEMeNT parameters, and in particular the mineralization rate constants, are assumed to take the same values for the different land use types. An exception is the protection coefficient, which takes a different value for cropland (h_c) compared to non-cultivated land (h_{nc}).

As in Van Meter et al. (2017), we start our simulations during the pre-industrial time (in year 1800 in this study). In the source zone, for the spatial units that initially have a non-cultivated land use (both agricultural permanent grassland and non-agricultural land), the active N store M_a^{prist} ($kg\ ha^{-1}$) and the protected N store M_p^{prist} ($kg\ ha^{-1}$) at the beginning of the simulations are set assuming steady state (pristine or pre-industrial) conditions:

$$M_a^{prist} = \frac{(1 - h_{nc}) Surplus_{non-agr}(t_0 \rightarrow t_1)}{k_a} \tag{S21}$$

$$M_p^{prist} = \frac{h_{nc} Surplus_{non-agr}(t_0 \rightarrow t_1)}{k_p} \tag{S22}$$

where h_{nc} (-) is the protection coefficient for non-cultivated land, $Surplus_{non-agr}(t_0 \rightarrow t_1)$ ($kg\ ha^{-1}\ yr^{-1}$) is the N surplus for non-agricultural land for the first simulation time

step, k_a (yr^{-1}) and k_p (yr^{-1}) are the rate constants of mineralization for the organic active and protected N store respectively.

The rate constant of mineralization for the organic protected store k_p is derived considering the following equation:

$$M_{s_{org}}^{prist} = M_a^{prist} + M_p^{prist} \tag{S23}$$

where $M_{s_{org}}^{prist}$ ($kg\ ha^{-1}$) is the total source zone organic N stock under pristine conditions, which is a calibrated parameter (see Table 1 in the main article). k_p is then calculated as:

$$k_p = \frac{h_{nc} Surplus_{non-agr}(t_0 \rightarrow t_1)}{M_{s_{org}}^{prist} - M_a^{prist}} \tag{S24}$$

In addition, in ELEMeNT, the protected N store for non-cultivated land (agricultural permanent grassland and non-agricultural land) cannot exceed its value under pristine conditions M_p^{prist}. It is thereby assumed that M_p^{prist} represents the maximum N storage capacity of the protected N store, and that no buildup of protected N can occur beyond the pristine conditions.

For source zone units that are cropland, the initial active N store M_a^{crop0} ($kg\ ha^{-1}$) is also determined assuming steady state conditions. The initial protected N store M_p^{crop0} ($kg\ ha^{-1}$) is set to 70% of its value under pristine conditions, assuming that a loss of 30% of protected N mass occurred after the conversion from non-cultivated to cultivated land (Van Meter et al., 2017):

$$M_a^{crop0} = \frac{(1 - h_c) Surplus_{crop}(t_0 \rightarrow t_1)}{k_a} \tag{S25}$$

$$M_p^{crop0} = 0.7 M_p^{prist} \tag{S26}$$

where $Surplus_{crop}(t_0 \rightarrow t_1)$ ($kg\ ha^{-1}\ yr^{-1}$) is the N surplus for cropland for the first simulation time step.

Subsurface Zone Initial State

As in previous applications of ELEMeNT, we set the initially subsurface N storage to zero. We use a long warm-up period (1800–1959) to attain a reasonable value of the storage, as explain in Section 4.1 in the main article.

Section S8. Derivation of the N Surplus for Cropland and Agricultural Permanent Grassland From the Total Agricultural N Surplus

At time t, the N surplus for agricultural permanent grassland, denoted as $Surplus_{mgra}(t)$ ($kg\ ha^{-1}\ yr^{-1}$), and for cropland, denoted as $Surplus_{crop}(t)$ ($kg\ ha^{-1}\ yr^{-1}$), can be estimated as follows:

$$Surplus_{crop}(t) = Surplus_{agr}(t)\frac{f_{agr}(t)}{f_{crop}(t) + r_{mgra-crop}f_{mgra}(t)} \tag{S27}$$

$$Surplus_{mgra}(t) = Surplus_{crop}(t)r_{mgra-crop} \tag{S28}$$

where $Surplus_{agr}(t)$ ($kg\ ha^{-1}\ yr^{-1}$) is the N surplus for agricultural areas, $r_{mgra-crop}$ (-) is the ratio of the N surplus for agricultural permanent grassland to the N surplus for cropland, $f_{agr}(t)$ (-) is the fraction of total agricultural areas, $f_{crop}(t)$ (-) is the fraction of cropland and $f_{mgra}(t)$ (-) is the fraction of agricultural permanent grassland. Equation S27 and S28 are derived from the following equality:

$$Surplus_{agr}(t)f_{agr}(t) = Surplus_{crop}(t)f_{crop}(t) + Surplus_{mgra}(t)f_{mgra}(t) \tag{S29}$$

Table S1. Datasets Used to Calculate the N Point Sources

Variable	Spatial resolution	Time period	Frequency	Source/reference
Population	5'x5'	1950–2016	decadal in 1950–2000; yearly in 2001-2016	HYDE V3.2.1, Klein Goldewijk et al. (2017)
Protein supply	Germany (NUTS0)	1950–2016	one value in 1950; yearly in 1961-2016	FAO (1951, 2021a, 2021b)
Food losses	Europe	2007	one value	FAO and SIK (2011)
Population connection to sewer and WWTPs	Germany	1950-1990	qualitative information	Seeger (1999)
	Germany (NUTS0)	1990–2016	variable	Eurostat (2021)
	Germany (NUTS2)	2010	one value	Eurostat (2016)
Observations of N loading from WWTPs	point data (WWTPs location)	2012-2016	1 value for each WWTP	Büttner (2020)

Note: NUTS: Nomenclature of territorial units for statistics. WWTP: Waste water treatments plants.

Table S2. Description of the Coefficients That Are Calibrated to Calculate the N Point Sources

Parameter	Description	Unit	Lower value	Upper value	References for the values
$f_{loss}^{protein}$	Fraction of protein supply lost at the distribution level	(-)	0.02	0.04	FAO and SIK (2011)
$f_{loss,house}^{protein}$	Fraction of protein supply lost at the household level	(-)	0.12	0.16	FAO and SIK (2011)
f_c^N	N content in protein	(kg N kg^{-1})	0.16	0.19	Mariotti et al. (2008), Morée et al. (2013)
$f_{loss,hum}^N$	Fraction of human N intake lost via sweat, hair and blood	(-)	0.02	0.04	best guess is 0.03 from Morée et al. (2013)
$f_{loss,sewer}^N$	Fraction of domestic N gross emissions collected by the sewer system leaked, settled, volatilized or degraded	(-)	0.05	0.15	best guess is 0.1 from Morée et al. (2013)
$f_{indus:house,1950}^N$	Ratio of industrial to domestic N gross emissions in 1950	(-)	0.6	0.9	best guess is 0.75 from Morée et al. (2013)
$f_{indus:house,2000}^N$	Ratio of industrial to domestic gross N emissions in 2000	(-)	0.1	0.25	best guess is 0.15 from Morée et al. (2013)
$f_{loss,indus}^N$	Fraction of industrial gross N emissions lost in stabilization ponds or through volatilization	(-)	0.2	0.4	best guess is 0.3 from Morée et al. (2013)
eff_{prim}^N	Efficiency of N removal for primary treatment	(-)	0.1	0.2	best guess is 0.1 from Van Drecht et al. (2009), Morée et al. (2013), and Van Puijenbroek et al. (2019)
eff_{sec}^N	Efficiency of N removal for secondary treatment	(-)	0.25	0.45	best guess is 0.35 from Van Drecht et al. (2009), Morée et al. (2013), and Van Puijenbroek et al. (2019)
eff_{ter}^N	Efficiency of N removal for tertiary treatment	(-)	0.7	0.94	best guess is 0.8 from Van Drecht et al. (2009), Morée et al. (2013), and Van Puijenbroek et al. (2019)

Table S3. Discharge (Q_{out}) at the Measuring Station, Source of the Daily Measurements, Time Periods for the Daily Measurements and Years for Which the Measurements Present Gaps For Each Subcatchment

Catchment outlet	Q_{out} station	Source of Q_{out} measurements	Time period of Q_{out} measurements	Years with gaps in Q_{out} measurements
Wahnhausen	Bonaforth	FGG	1978-1983	1984, 1999-2000
Letzter Heller	Letzter Heller	GRDC	1941-2015	-
Hemeln	Hann. Münden	GRDC	1831-2015	-
Hessisch Oldendorf	Hameln	FGG	1975-2015	1981-1982
Porta	Porta	GRDC	1936-2015	-
Petershagen	Petershagen	FGG	1986-2015	-
Drakenburg	Drakenburg	FGG	1979-2015	1984-1985
Verden	Westen	FGG	1970-1983	1984-1989, 2001-2005
Hemelingen	Intschede	GRDC	1857-2015	-

Note: Discharge measurements are obtained from the River Basin Commission Weser (FGG Weser, 2021) and from the (Global Runoff Data Centre, 2021). We report here the years that present long gaps (i.e., longer than five days). Short gaps (i.e., less or equal to five days) are filled using linear interpolation.

Table S4. Performance of mHM Discharge Simulations at the Outlet of the Subcatchments

Station	Medium-term simulations (daily)				Long-term simulations (annual)			
	Number days	*PBIAS* (%)	R^2 (-)	*NSE* (-)	Number years	*PBIAS* (%)	R^2 (-)	*NSE* (-)
Bonaforth	13,179	1.8	0.87	0.87	35	30	0.85	0.83
Letzter Heller	24,106	4.1	0.89	0.89	75	26	0.71	0.68
Hann. Münden	24,106	0.29	0.91	0.91	185	28	0.66	0.64
Hameln	14,671	4.0	0.93	0.93	39	26	0.83	0.79
Porta	24,106	2.7	0.92	0.91	80	22	0.83	0.78
Drakenburg	13,210	6.8	0.93	0.93	35	29	0.85	0.82
Westen	10,500	4.2	0.80	0.74	25	63	0.83	0.76
Intschede	24,106	1.3	0.90	0.89	159	34	0.71	0.68

Note: The medium-term simulations refer to the simulations performed over the period 1950–2015, using the mHM set-up described in Zink et al. (2017). The long-term simulations refer to the simulations performed over the period 1766–2015, using the mHM setup described in Hanel et al. (2018). We note that long-term simulations are available at Intschede only, and they are used to reconstruct the long-term discharge for all stations (these simulations provide the long-term trend). Performance metrics are evaluated at daily timescale for the medium-term simulations, and at annual timescale for the long-term simulations. The table reports the number of days or years that are used to calculate the performance metrics. *PBIAS* is the percent bias (absolute value), R^2 is the coefficient of determination and NSE is the Nash–Sutcliffe efficiency. **Reported *PBIAS* values are calculated before bias correction**. We perform bias correction using a multiplicative factor, so that *PBIAS* is equal to 0 after bias correction. **Reported *NSE* values are calculated after bias correction.** We verify that the mHM simulations perform well, with values of *NSE* always higher than 0.64, and values of R^2 always higher than 0.66. For the medium-term simulations of Zink et al. (2017), the performance is particulary high, with values of *NSE* always higher than 0.74 and values of R^2 always higher than 0.80.

Table S5. Range of Observed Soil N Content in 2009 in the Horizon 0-100 cm (M_s) for the Eight Subcatchments

Catchment outlet	Observed M_s ($kg\ ha^{-1}$)	
	Lower limit	Upper limit
Wahnhausen	16,070	25,712
Letzter Heller	14,981	23,969
Hemeln	15,555	24,888
Hessisch Oldendorf	15,419	24,671
Porta	15,239	24,382
Petershagen	15,239	24,382
Drakenburg	14,895	23,831
Verden	12,618	20,189
Hemelingen	13,930	22,287

Note: Values are obtained by combining the N content and bulk density in the topsoil (0–20 cm) from the LUCAS dataset (Ballabio et al., 2016, 2019), and the ratio of total soil N content (0–100 cm) to topsoil N content (0–20 cm), which we estimate to be between 2.5 (lower limit) and 4 (upper limit) from Table 4 in Batjes (1996) We note that the study of Batjes (1996) reports the average soil N content for the horizons 0-30 cm and 0-100 cm for a range of soil types. We estimate the value for a depth of 0–20 cm from the 0–30 cm values of Batjes (1996) assuming that, in the first 30 cm of soil, the N content is proportional to the depth.

Table S6. Number of Behavioural Simulations Obtained After Application of the Seven Soft Rules For Each Subcatchment

Catchment outlet	Number of behavioural simulations
Wahnhausen	2076
Letzter Heller	676
Hemeln	1304
Hessisch Oldendorf	1503
Porta	1618
Drakenburg	1726
Verden	679
Hemelingen	1475

Note: The methodology used to obtain the behavioural simulations is described in Section 4.1 of the main article.

Table S7. Prior and Posterior Median and 95% Confidence Interval of the ELEMeNT Parameter Values for the Eight Subcatchments

Parameter		Prior	Wahnhausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Drakenburg	Verden	Hemelingen
$M_{s_{org}}^{prist}$	Q_{50}	22493	22932	21623	22494	22135	21808	21445	18269	20020
$(kg\ ha^{-1})$	$Q_{2.5}$	10621	17786	16559	17167	16966	17014	16380	14234	15612
	$Q_{97.5}$	34373	28200	26367	27365	26992	26906	26345	22609	24695
h_c	Q_{50}	0.27	0.23	0.24	0.22	0.27	0.24	0.27	0.25	0.25
(-)	$Q_{2.5}$	0.11	0.11	0.11	0.10	0.11	0.11	0.11	0.11	0.11
	$Q_{97.5}$	0.48	0.47	0.47	0.46	0.48	0.46	0.48	0.49	0.48
h_{nc}	Q_{50}	0.54	0.52	0.54	0.49	0.54	0.52	0.54	0.52	0.54
(-)	$Q_{2.5}$	0.28	0.27	0.27	0.27	0.28	0.28	0.28	0.27	0.27
	$Q_{97.5}$	0.74	0.74	0.74	0.73	0.74	0.74	0.74	0.74	0.74
k_a	Q_{50}	0.40	0.50	0.33	0.39	0.38	0.48	0.37	0.44	0.49
(yr^{-1})	$Q_{2.5}$	0.07	0.18	0.10	0.14	0.10	0.18	0.09	0.13	0.17
	$Q_{97.5}$	0.73	0.74	0.72	0.73	0.73	0.74	0.72	0.73	0.74
V_s	Q_{50}	300	306	375	349	301	304	295	347	316
(mm)	$Q_{2.5}$	110	111	137	114	109	113	109	137	112
	$Q_{97.5}$	490	492	496	492	491	486	491	488	492
λ_s	Q_{50}	0.55	0.34	0.28	0.30	0.32	0.31	0.33	0.28	0.33
(yr^{-1})	$Q_{2.5}$	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
	$Q_{97.5}$	0.98	0.71	0.63	0.63	0.65	0.63	0.67	0.60	0.66
λ_{sub}	Q_{50}	0.15	0.17	0.09	0.11	0.07	0.14	0.06	0.16	0.16
(yr^{-1})	$Q_{2.5}$	0.02	0.05	0.03	0.04	0.02	0.03	0.02	0.05	0.04
	$Q_{97.5}$	0.29	0.29	0.27	0.26	0.24	0.29	0.25	0.29	0.29
μ_{sub}	Q_{50}	26	10	17	14	16	8	19	11	8
(yr)	$Q_{2.5}$	3	3	4	4	4	2	3	3	2
	$Q_{97.5}$	49	29	36	34	40	23	44	31	24
R	Q_{50}	0.16	0.11	0.16	0.12	0.14	0.12	0.14	0.15	0.12
(-)	$Q_{2.5}$	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.02	0.01
	$Q_{97.5}$	0.29	0.29	0.29	0.29	0.29	0.28	0.29	0.29	0.28

Note: The prior distributions are the distributions in the initial sample of size 100,000 and are the same for all subcatchments. The posterior distributions correspond to the distributions in the behavioural sample obtained after application of the seven soft rules. The values correspond to the median (Q_{50}) and the lower and upper bound of the 95% confidence interval ($Q_{2.5}$ and $Q_{97.5}$ respectively).

Table S8. Percentage Change in the Posterior Values Compared to the Prior Values of the Median and 95% Confidence Interval for the ELEMeNT Parameters for the Eight Subcatchments

Parameter		Wahnhausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Drakenburg	Verden	Hemelingen
$M_{s_{org}}^{prist}$	ΔQ_{50}	2	-4	0	-2	-3	-5	-19	-11
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-56	-59	-57	-58	-58	-58	-65	-62
h_c	ΔQ_{50}	-13	-11	-19	1	-11	0	-6	-6
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-4	-3	-5	0	-6	-1	0	-1
h_{nc}	ΔQ_{50}	-3	1	-10	1	-3	1	-3	-1
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	1	1	-0	-0	-1	-1	1	0
k_a	ΔQ_{50}	26	-18	-2	-6	19	-8	11	22
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-16	-7	-10	-6	-15	-5	-10	-15
V_s	ΔQ_{50}	2	25	16	0	1	-2	16	5
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	0	-6	-0	1	-2	0	-7	-0
λ_s	ΔQ_{50}	-37	-49	-46	-41	-44	-40	-49	-39
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-30	-39	-39	-37	-39	-35	-42	-35
λ_{sub}	ΔQ_{50}	11	-43	-31	-54	-8	-59	1	2
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-13	-12	-19	-21	-8	-16	-12	-8
μ_{sub}	ΔQ_{50}	-63	-33	-47	-37	-70	-27	-59	-68
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-41	-31	-34	-20	-56	-11	-39	-53
R	ΔQ_{50}	-27	2	-20	-10	-22	-9	-5	-23
(%)	$\Delta(Q_{97.5} - Q_{2.5})$	-2	-0	-2	-1	-2	-1	-1	-2

Note: The prior distributions are the distributions in the initial sample of size 100,000. The posterior distributions correspond to the distributions in the behavioural sample obtained after application of the seven soft rules. The values correspond to the percentage change in the median (ΔQ_{50}) and the 95% confidence interval ($\Delta(Q_{97.5} - Q_{2.5})$).

Table S9. Posterior Median and 95% Confidence Interval of the Mean Transfer Times for Organic Active and Protected N Stores

Parameter		Wahnhausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Drakenburg	Verden	Hemelingen
t_a (yr)	Q_{50}	2.0	3.0	2.5	2.7	2.1	2.7	2.3	2.0
	$Q_{2.5}$	1.4	1.4	1.4	1.4	1.4	1.4	1.4	1.4
	$Q_{97.5}$	5.6	9.7	7.3	9.9	5.7	11.3	7.6	5.9
t_p (yr)	Q_{50}	2597	2336	2646	2392	2454	2375	2079	2268
	$Q_{2.5}$	1579	1444	1527	1471	1529	1486	1323	1402
	$Q_{97.5}$	5376	4831	5331	4841	5025	4739	4203	4505

Note: t_a: mean transfer time for organic active N store, which is equal to the inverse of the mineralization rate constant for organic active N store $1/k_a$; t_p: mean transfer time for organic protected N store, which is equal to the inverse of the mineralization rate constant for organic protected N store $1/k_p$. k_p is calculated from Equation S24 and its CDFs are represented in Figure S28. The values correspond to the median (Q_{50}) and the lower and upper bound of the 95% confidence interval ($Q_{2.5}$ and $Q_{97.5}$ respectively).

Table S10. Matrix of Pearson Correlation Coefficients Between Pairs of ELEMeNT Parameters in the Behavioural Sample for the **Hemelingen** Subcatchment

	$M_{s_{org}}^{prist}$	h_c	h_{nc}	k_a	V_s	λ_s	λ_{sub}	μ_{sub}	R
$M_{s_{org}}^{prist}$	1.0	**-0.12**	0.0	-0.02	**-0.06**	0.01	0.01	0.03	-0.02
h_c	**-0.12**	1.0	**0.27**	-0.01	0.05	**-0.06**	-0.02	**-0.15**	-0.02
h_{nc}	0.0	**0.27**	1.0	-0.02	-0.01	-0.01	-0.02	-0.05	-0.01
k_a	-0.02	-0.01	-0.02	1.0	-0.01	-0.02	**-0.13**	**0.18**	0.01
V_s	**-0.06**	0.05	-0.01	-0.01	1.0	**-0.09**	0.03	**-0.09**	0.01
λ_s	0.01	**-0.06**	-0.01	-0.02	**-0.09**	1.0	**-0.36**	**-0.54**	**-0.08**
λ_{sub}	0.01	-0.02	-0.02	**-0.13**	0.03	**-0.36**	1.0	**-0.32**	-0.02
μ_{sub}	0.03	**-0.15**	-0.05	**0.18**	**-0.09**	**-0.54**	**-0.32**	1.0	**-0.28**
R	-0.02	-0.02	-0.01	0.01	0.01	**-0.08**	-0.02	**-0.28**	1.0

Note: We highlight significant values in bold (significance level equal to 0.05). We observe particularly high values (higher than 0.25) of the correlation coefficient between the denitrification rate constant in the source zone λ_s and the denitrification rate constant in the subsurface λ_{sub} (-0.36), between λ_s and the mean travel time in the subsurface μ_{sub} (-0.54), between μ_{sub} and λ_{sub} (-0.32), between μ_{sub} and the fraction of in-stream N removal R (-0.28), and between the protection coefficient for cultivated land h_c and non-cultivated land h_{nc} (0.27). This indicates that these parameters are interacting. We note that the analysis of the pairwise correlation coefficient only allows to detect two-way interactions and not higher order interactions We also note that the two protection coefficients are not sampled independently in the prior parameter sample. In fact, we discard the parameter sets that do not meet the condition $h_c < h_{nc}$, as explained in Section 4.1 in the main article. Therefore, the relatively high values of the correlation coefficient between h_c and h_{nc} could be due to the initial sampling strategy and not to the conditions imposed by the soft rules.

Table S11. Simulated Contribution (in %) of the Denitrification in the Source Zone to the Total Denitrification Aggregated Over the Source Zone and the Subsurface for the Eight Subcatchments

Time period	Stat.	Wahnhausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Drakenburg	Verden	Hemelingen
1960–2015	Q_{50}	50.5	46.8	46.4	51.6	50.5	53.3	48.6	53.8
	$Q_{2.5}$	16.3	17.4	17.0	18.2	17.6	18.9	19.5	17.8
	$Q_{97.5}$	93.5	92.9	91.8	93.4	92.1	94.2	89.5	94.2

The table reports the median (Q_{50}), the lower bound ($Q_{2.5}$, 2.5^{th} percentile) and upper bound ($Q_{97.5}$, 97.5^{th} percentile) values in the **behavioural** simulation ensemble for the period 1960–2015. The table shows that the range of variation (between lower and upper bounds) is very large for all subcatchments. This means that there are large uncertainties regarding the location (source zone or subsurface) of denitrification.

Table S12. Source Zone Storage: Initial Condition in 1960 and Change for the Period 1960–2015

Variable	Stat	Wahn-hausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Draken-burg	Verden	Hemelingen
M_s^{init}	Q_{50}	20,468	19,176	19,981	19,666	19,244	18,911	16,020	17,487
(kg ha^{-1})	$Q_{2.5}$	15,910	14,648	15,302	15,192	15,040	14,530	12,499	13,678
	$Q_{97.5}$	25,039	23,330	24,187	23,901	23,666	23,131	19,698	21,564
ΔM_s	Q_{50}	2.5	2.5	2.4	2.6	2.6	3.1	2.0	2.5
(% M_s^{init})	$Q_{2.5}$	1.3	1.3	1.3	1.3	1.4	1.5	0.5	1.2
	$Q_{97.5}$	4.5	4.7	4.4	4.9	4.6	5.6	4.4	5.1
ΔM_s	Q_{50}	504	480	473	526	485	584	326	448
(kg ha^{-1})	$Q_{2.5}$	290	270	283	278	283	306	80	229
	$Q_{97.5}$	828	764	786	803	807	911	642	781
ΔM_p	Q_{50}	502	452	461	500	471	549	277	416
(kg ha^{-1})	$Q_{2.5}$	284	244	270	254	267	269	43	203
	$Q_{97.5}$	827	733	773	785	793	880	594	756
ΔM_a	Q_{50}	3	13	7	12	10	20	31	21
(kg ha^{-1})	$Q_{2.5}$	-1	1	1	4	4	8	18	12
	$Q_{97.5}$	21	115	55	104	50	150	128	74
ΔM_{in}	Q_{50}	-2	1	0	0	0	-2	7	2
(kg ha^{-1})	$Q_{2.5}$	-9	-6	-6	-7	-6	-9	-8	-5
	$Q_{97.5}$	3	8	6	8	7	6	23	16

Notations: The table reports the median (Q_{50}), the lower bound ($Q_{2.5}$, 2.5^{th} percentile) and upper bound ($Q_{97.5}$, 97.5^{th} percentile) values in the **behavioural** simulation ensemble. M_s^{init} is the initial condition for the source zone storage in 1960; ΔM_s, ΔM_p, ΔM_a, and ΔM_{in} are the change in the total, protected, active and inorganic storage in the source zone, respectively, for the period 1960–2015.
($\Delta M_s = \Delta M_p + \Delta M_a + \Delta M_{in}$)

Table S13. Simulated Contribution (in %) of the N Point Sources to the Total In-stream N Loading at the Catchment Outlet for the Eight Subcatchments

Time period	Stat.	Wahnhausen	Letzter Heller	Hemeln	Hessisch Oldendorf	Porta	Drakenburg	Verden	Hemelingen
1960–2015	Q_{50}	28.2	20.6	23.6	21.6	23.5	25.8	34.4	28.7
	$Q_{2.5}$	21.0	15.8	18.0	16.4	17.6	19.7	28.9	22.6
	$Q_{97.5}$	35.8	26.2	30.2	27.3	29.6	32.7	39.4	35.6
2006–2015	Q_{50}	19.6	14.3	16.5	13.8	14.2	14.5	20.5	17.1
	$Q_{2.5}$	14.3	10.6	12.1	10.0	10.3	10.6	16.6	12.9
	$Q_{97.5}$	26.5	18.9	21.8	18.2	18.7	19.5	24.7	22.7
1995–2002	Q_{50}	17.2	13.2	14.6	12.2	13.2	13.5	18.2	15.4
	$Q_{2.5}$	12.5	9.7	10.8	8.9	9.7	9.9	14.7	11.8
	$Q_{97.5}$	23.0	17.5	19.2	16.0	17.4	17.9	21.8	20.3

The table reports the median (Q_{50}), the lower bound ($Q_{2.5}$, 2.5^{th} percentile) and upper bound ($Q_{97.5}$, 97.5^{th} percentile) values in the **behavioural** simulation ensemble. The contribution of the N point sources is calculated as the ratio of the cumulative simulated in-stream N loading originating from N point sources ($J_{out_{ps}}$, which is equal to the N point sources input, from which we deduct the in-stream N removal $J_{rem_{ps}}$) to the cumulative simulated total in-stream N loading at the catchment outlet (J_{out}) for three time periods, 1950–2016, 1995–2002 (to be compared with the results for the study of Grizzetti et al., 2008) and 2006–2015.

Table S14. Number of Behavioural Simulations for Hemelingen for the 27 N Surplus Scenarios and the Ten N Point Source Realizations

N surplus scenario	scenario	$f_{surplus}$	$r_{mgra-crop}$	r_{warm}	N point source realization									
					1	2	3	4	5	6	7	8	9	10
	1	0.8	0.5	0.25	882	551	726	778	789	831	872	911	971	1058
	2	0.8	0.5	0.5	886	531	710	767	779	826	872	910	965	1058
	3	0.8	0.5	0.75	1095	722	930	985	1006	1045	1082	1128	1176	1268
	4	0.8	1	0.25	1098	726	943	1002	1006	1040	1088	1108	1163	1270
	5	0.8	1	0.5	1170	794	1011	1066	1078	1109	1153	1191	1234	1351
	6	0.8	1	0.75	1179	793	1018	1075	1085	1119	1166	1199	1242	1359
	7	0.8	1.5	0.25	1220	860	1050	1125	1130	1163	1205	1232	1289	1398
	8	0.8	1.5	0.5	1224	860	1058	1133	1143	1167	1210	1238	1297	1398
	9	0.8	1.5	0.75	1223	861	1061	1131	1139	1177	1214	1240	1303	1404
	10	1	0.5	0.25	1146	755	994	1064	1068	1104	1131	1169	1243	1362
	11	1	0.5	0.5	1157	761	996	1059	1069	1099	1147	1187	1249	1357
	12	1	0.5	0.75	1383	980	1201	1256	1267	1325	1368	1415	1477	1611
	13	1	1	0.25	1390	983	1200	1273	1283	1327	1373	1406	1470	1595
	14	1	1	0.5	**1475**	1058	1296	1361	1368	1413	1462	1513	1575	1716
	15	1	1	0.75	1483	1066	1304	1362	1371	1418	1471	1520	1585	1718
	16	1	1.5	0.25	1537	1117	1348	1422	1440	1480	1523	1566	1625	1768
	17	1	1.5	0.5	1552	1130	1364	1445	1451	1489	1536	1586	1642	1792
	18	1	1.5	0.75	1561	1133	1373	1441	1451	1498	1540	1597	1648	1806
	19	1.2	0.5	0.25	1357	888	1128	1219	1229	1293	1345	1394	1466	1647
	20	1.2	0.5	0.5	1370	872	1147	1229	1246	1297	1357	1405	1500	1658
	21	1.2	0.5	0.75	1667	1131	1424	1524	1534	1582	1652	1716	1790	1947
	22	1.2	1	0.25	1690	1135	1449	1544	1550	1600	1669	1734	1806	1959
	23	1.2	1	0.5	1794	1244	1547	1634	1647	1704	1773	1827	1918	2072
	24	1.2	1	0.75	1791	1232	1537	1623	1631	1694	1764	1816	1905	2058
	25	1.2	1.5	0.25	1854	1292	1605	1690	1706	1759	1825	1894	1991	2141
	26	1.2	1.5	0.5	1863	1304	1614	1700	1715	1778	1840	1912	2003	2146
	27	1.2	1.5	0.75	1864	1292	1605	1692	1702	1768	1834	1894	2001	2145

Note: In red and bold is the baseline scenario chosen for the analyses ($f_{surplus}$ = 1, $r_{mgra-crop}$ = 1, r_{warm}=0.5 and N point source realization 1). $f_{surplus}$: N surplus multiplier; $r_{mgra-crop}$ (-): ratio of N surplus for agricultural permanent grassland to N surplus for cropland; r_{warm} (-): ratio of the value of the agricultural N surplus in 1850 to the value in 1950. The number of behavioural simulations vary between 531 and 2146. Therefore, we identify a large number of behavioural simulations for all combinations of N surplus scenario and N point source realization.

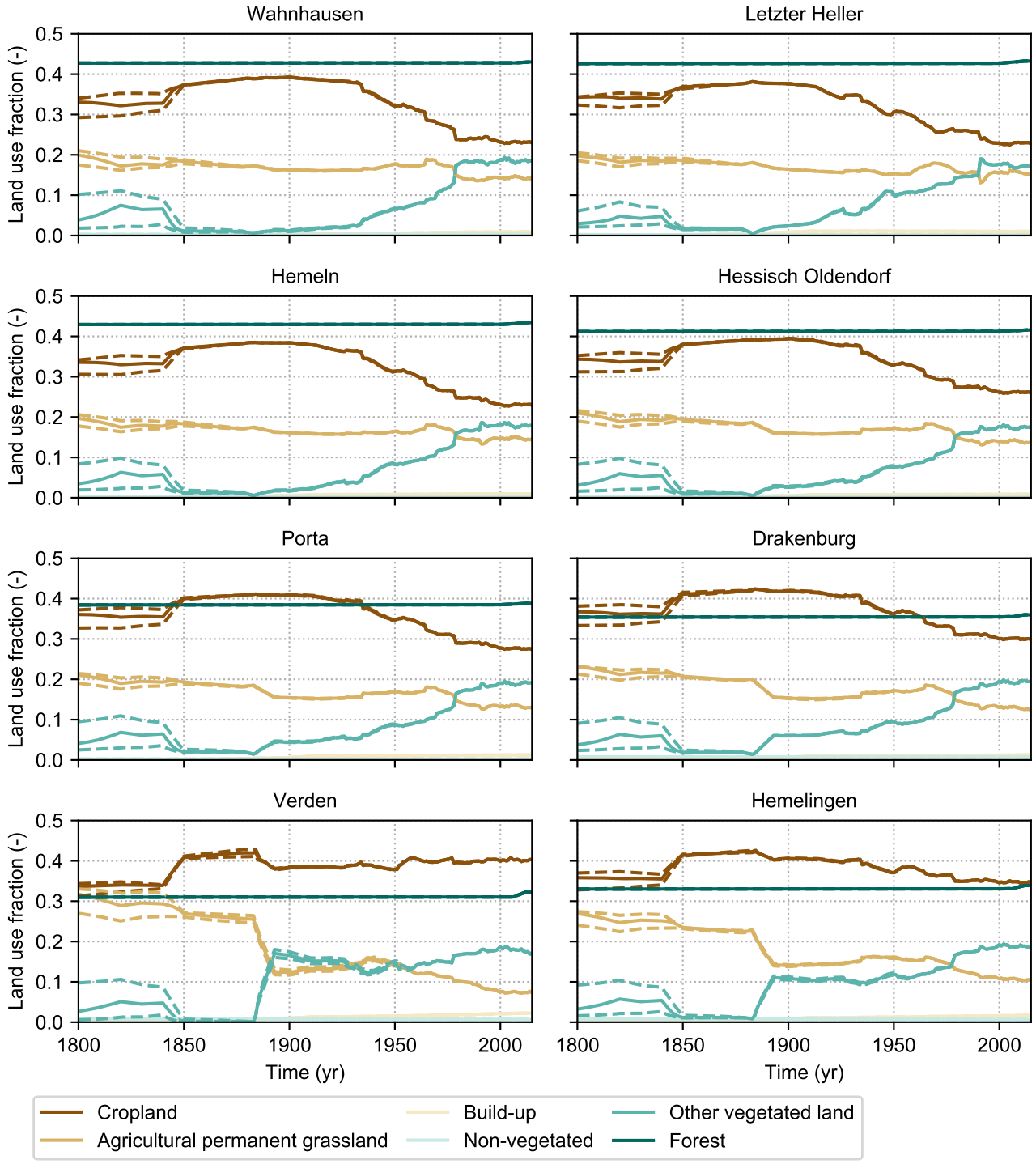

Figure S1. Time series of the land use fractions, i.e., cropland, agricultural permanent grassland, built-up areas, non-vegetated areas, other vegetated land (which includes in particular natural grassland, urban parks, and green areas in discontinuous urban fabric) and forest for the period 1800–2015 for the eight subcatchments. The solid lines are obtained using the baseline scenario for the HYDE dataset, while the dotted lines are obtained using the upper and lower scenarios.

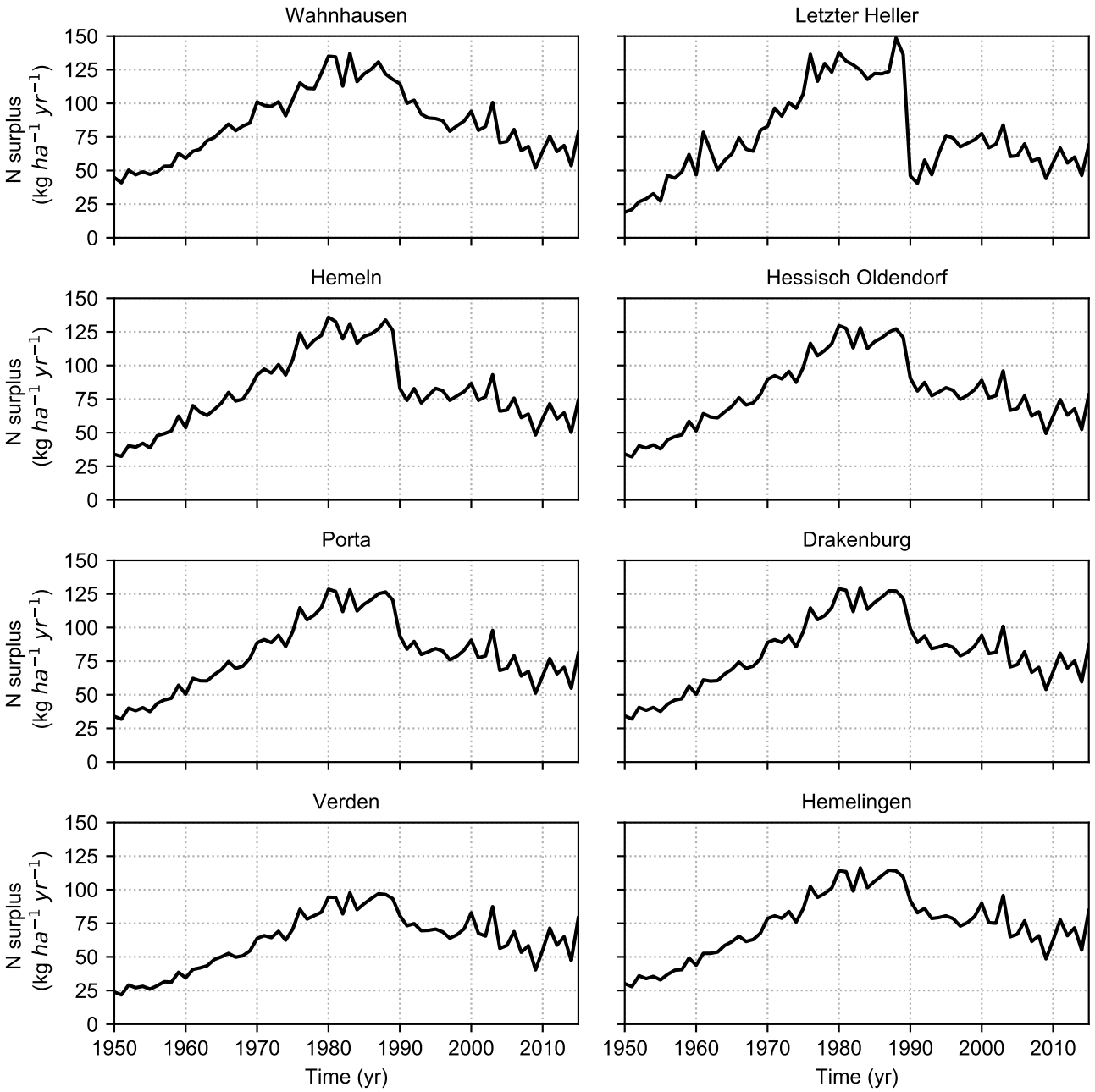

Figure S2. Time series of the agricultural N surplus **(per unit of agricultural area)** for the period 1950–2015 for the eight subcatchments.

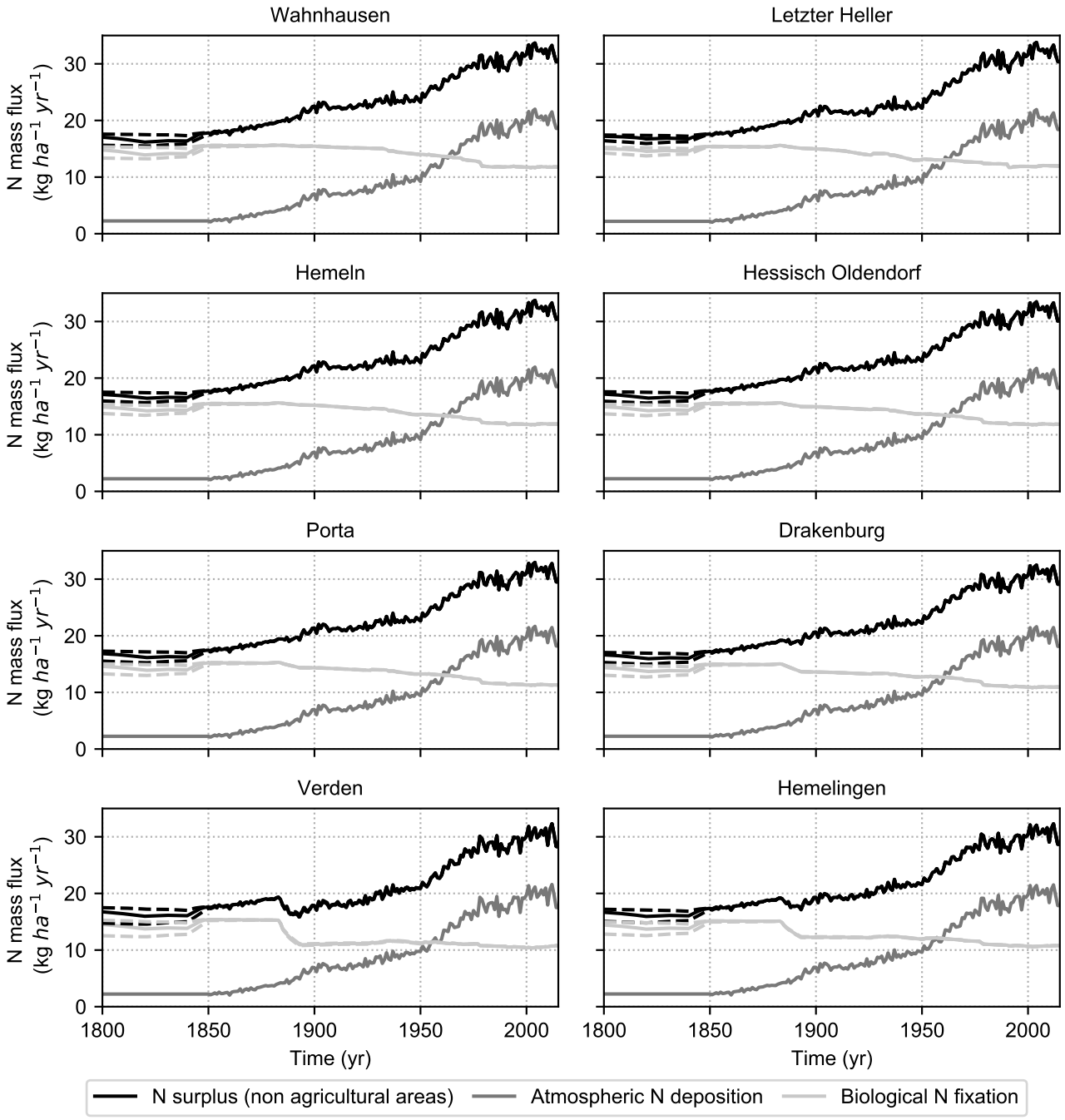

Figure S3. Time series of the total N surplus for non-agricultural areas and its components i.e., atmospheric N deposition and biological N fixation **(per unit of non-agricultural areas)** for the period 1800–2015 for the eight subcatchments. For biological N fixation, the solid lines are obtained using the land use data corresponding to the baseline scenario of the HYDE dataset, while the dotted lines are obtained using the land use data corresponding to the upper and lower scenarios of the HYDE dataset. We also refer to Section S5 for further explanation. We only consider the baseline scenario to force the ELEMeNT model, given the small uncertainties.

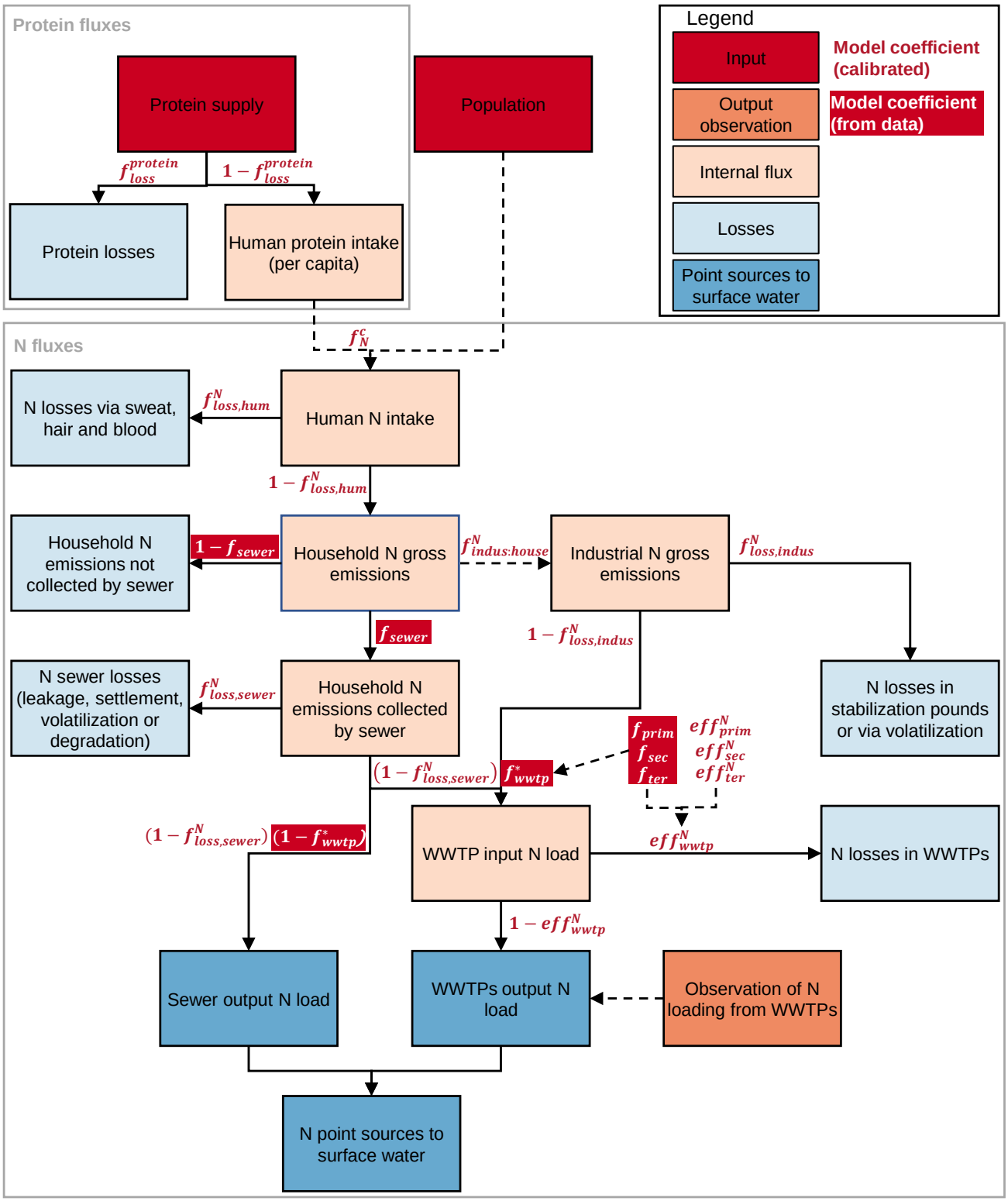

Figure S4. Flow chart depicting the procedure for the calculation of the N point sources. Details on the datasets used are reported in Table S1, and calibrated model coefficients are described in Table S2. WWTP: Waste Water Treatment Plant.

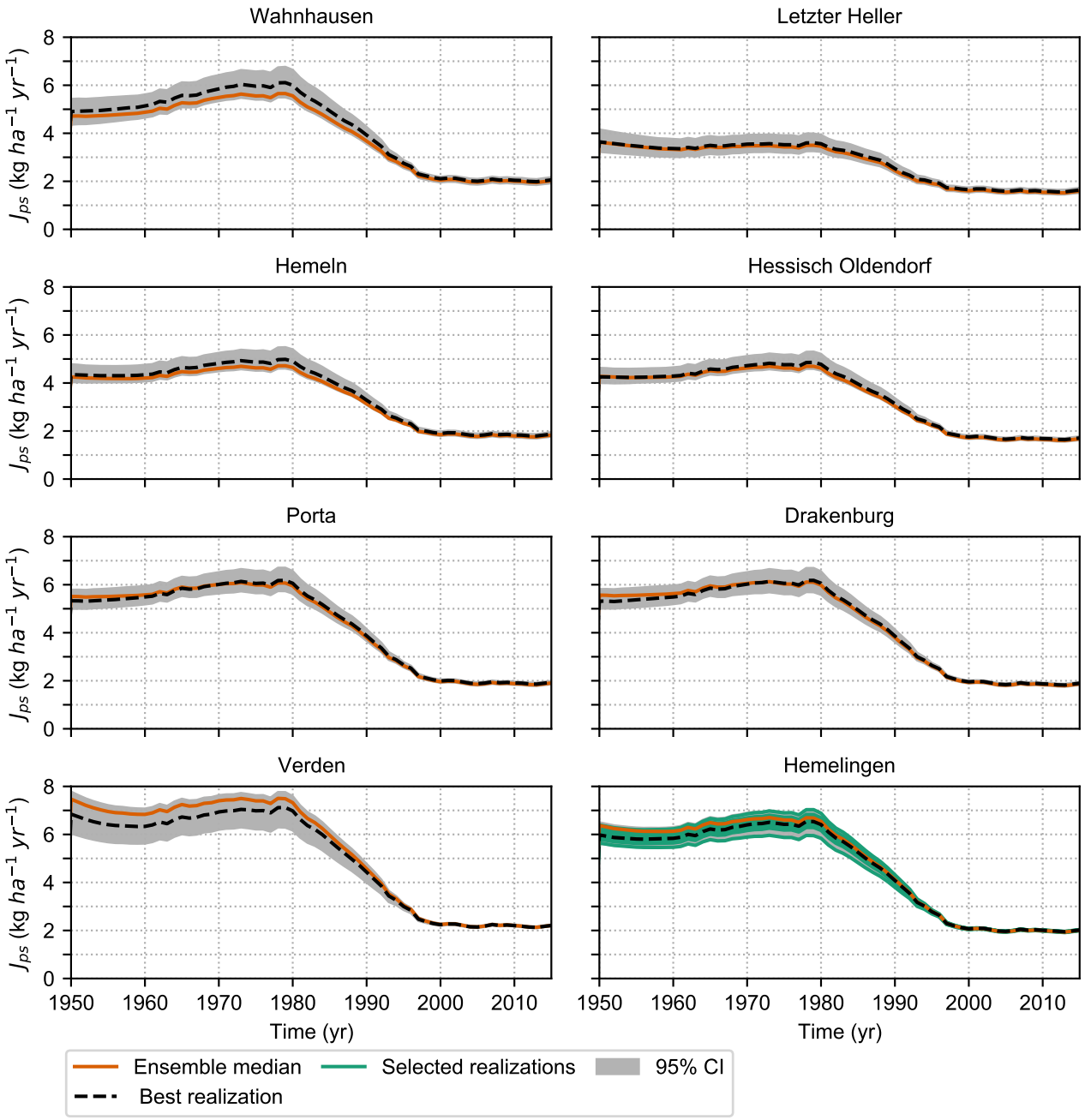

Figure S5. Time series of the N point sources for the period 1950–2015 for the eight subcatchments. The dashed black lines represent the baseline realization that we select to estimate the model parameters for the eight subcatchments (Section 4.1). This is the realization that presents the smallest error with respect to the observation dataset of N loading from waste water treatment plants of Büttner (2020). The green lines in the panel for Hemelingen represent the nine additional realizations that we select to perform the sensitivity analysis (Section 4.2).

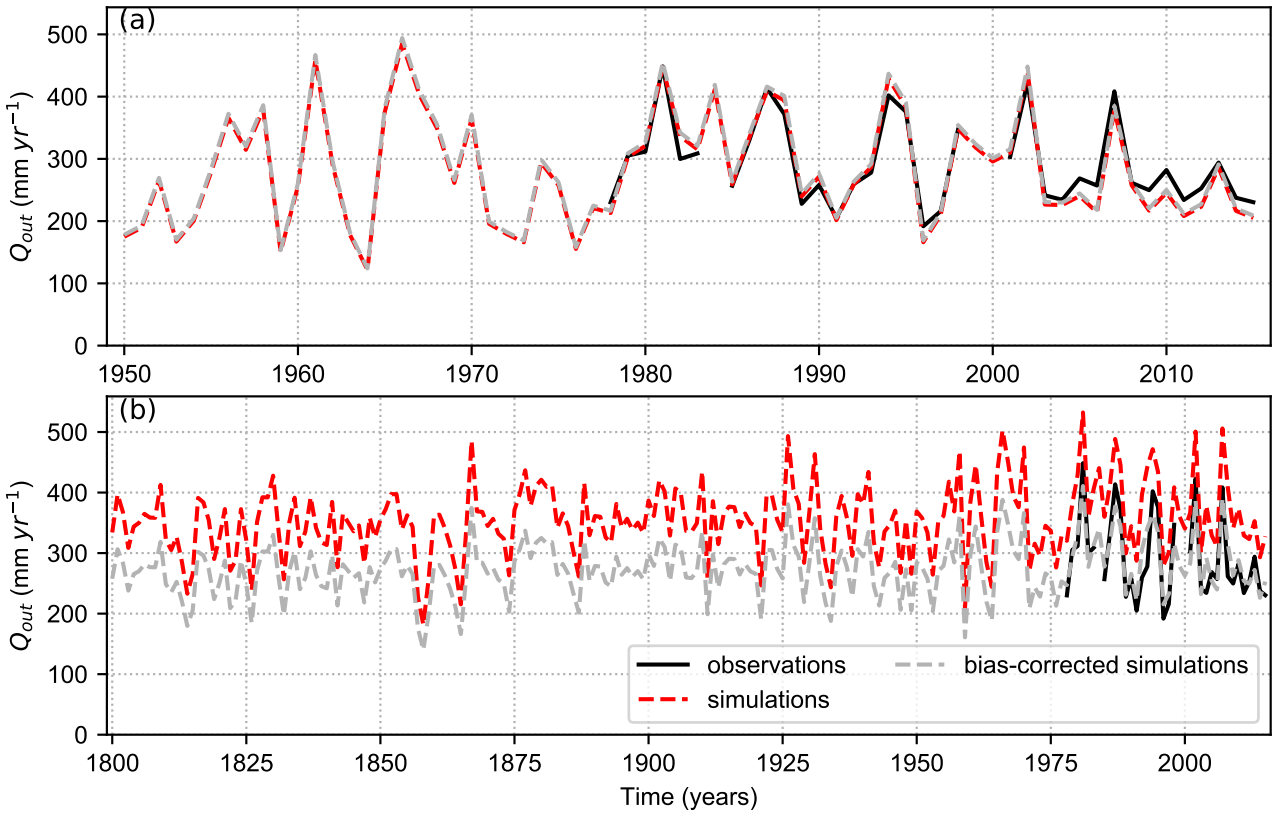

Figure S6. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Bonaforth** station (Wahnhausen subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

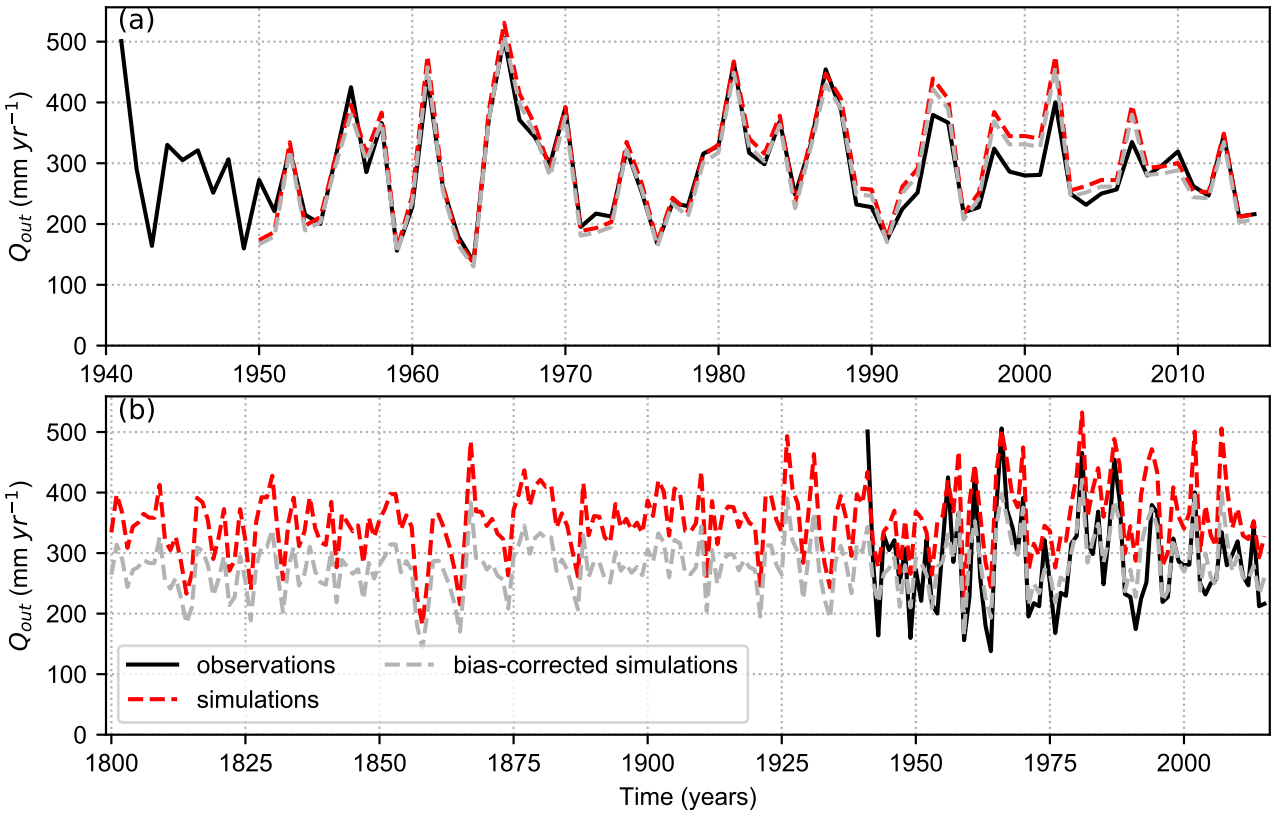

Figure S7. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Letzter Heller** station (Letzter Heller subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

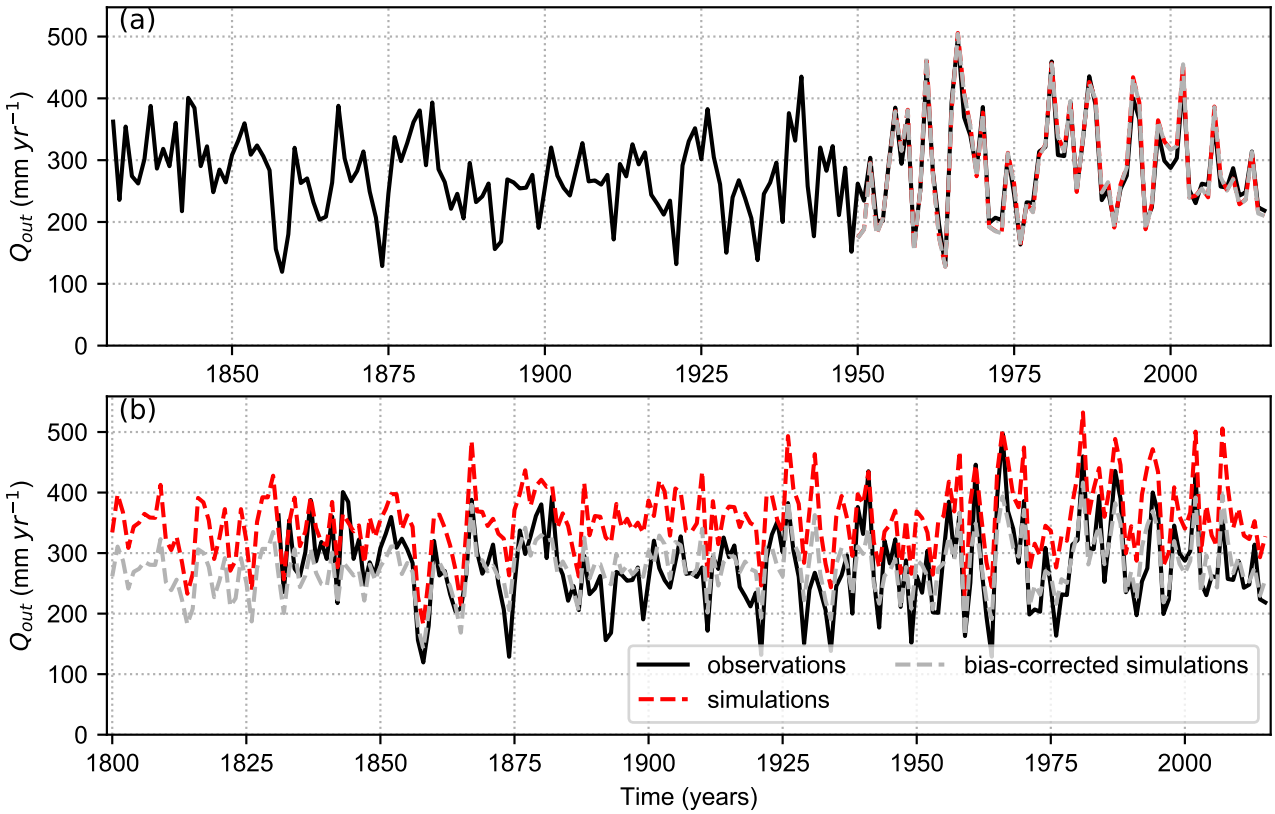

Figure S8. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Hann. Münden** station (Hemeln subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

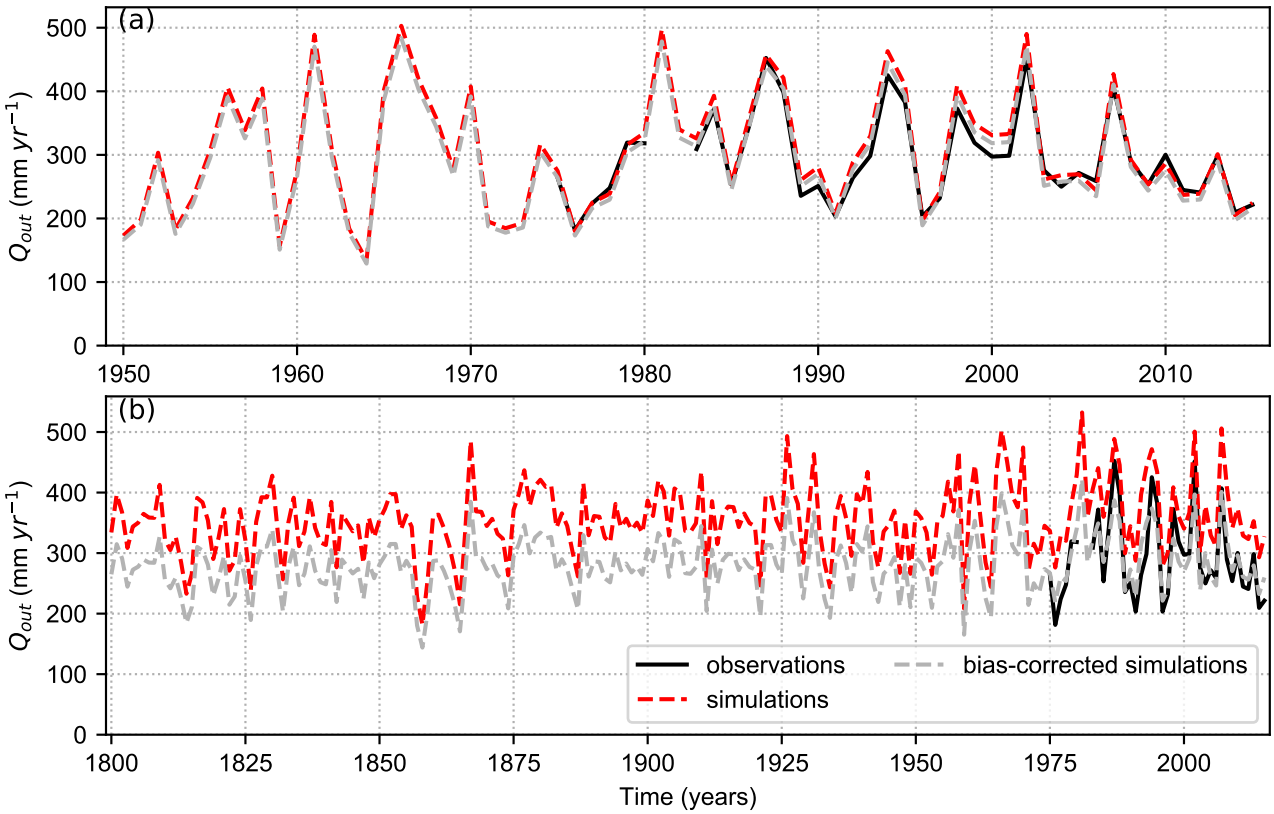

Figure S9. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Hameln** station (Hessisch Oldendor subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

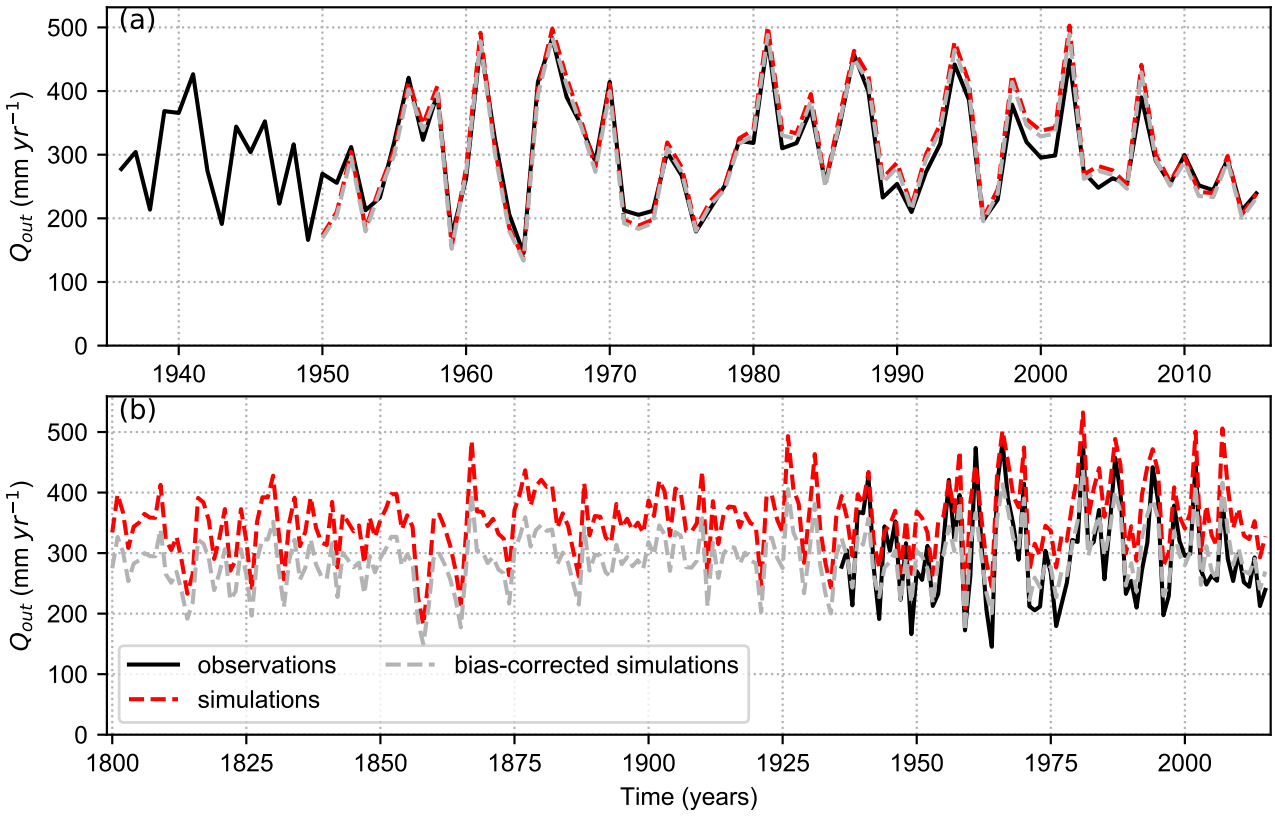

Figure S10. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Porta** station (Porta subcatchment): (a) medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

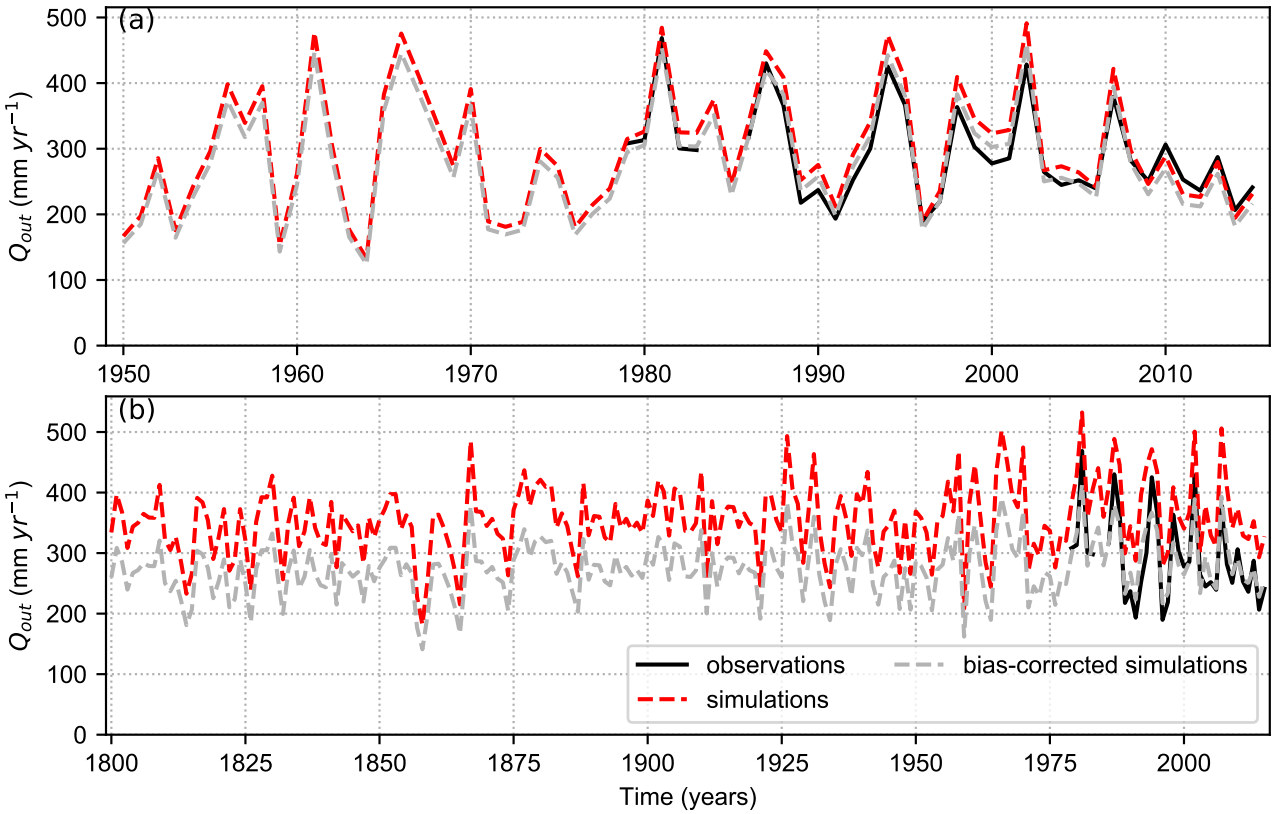

Figure S11. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Drakenburg** station (Drakenburg subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

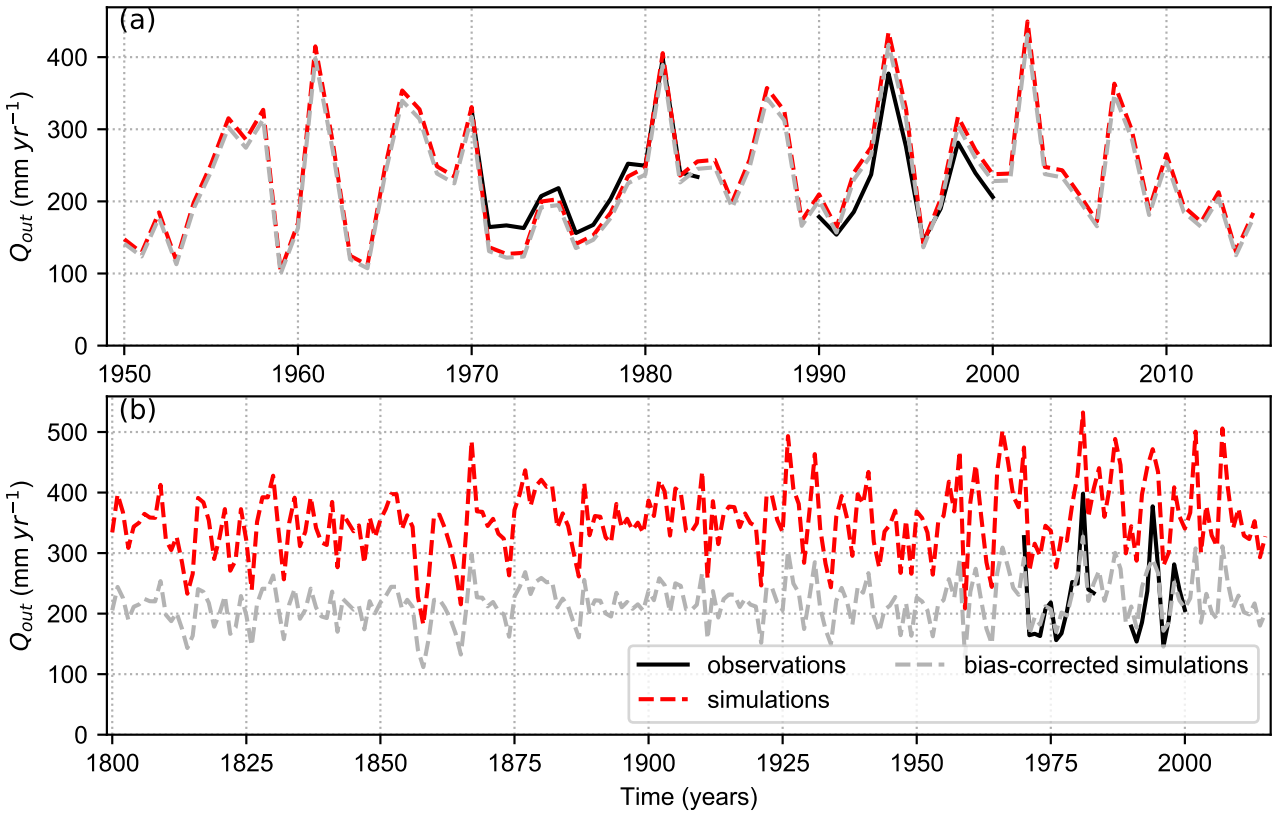

Figure S12. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Westen** station (Verden subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

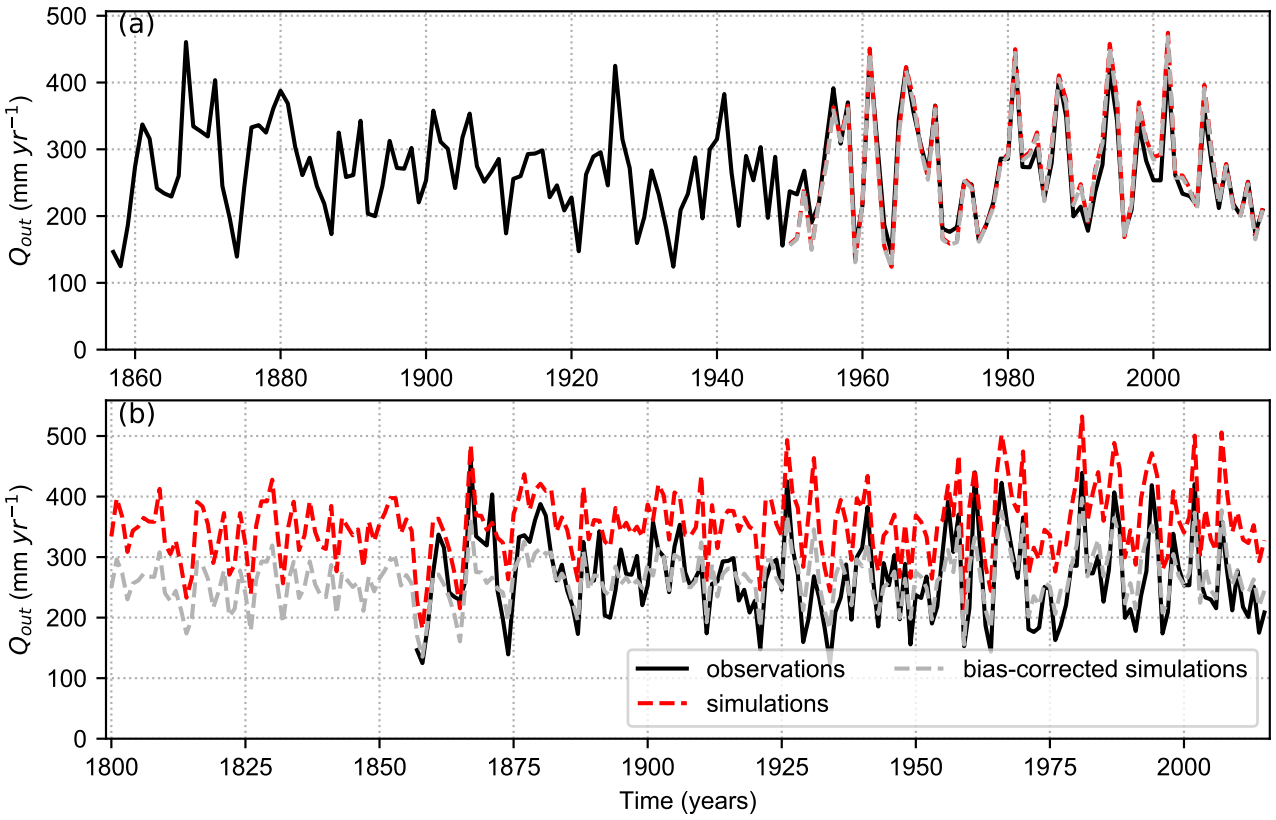

Figure S13. Observed, simulated and bias-corrected simulated annual discharge Q_{out} for the **Intschede** station (Hemelingen subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

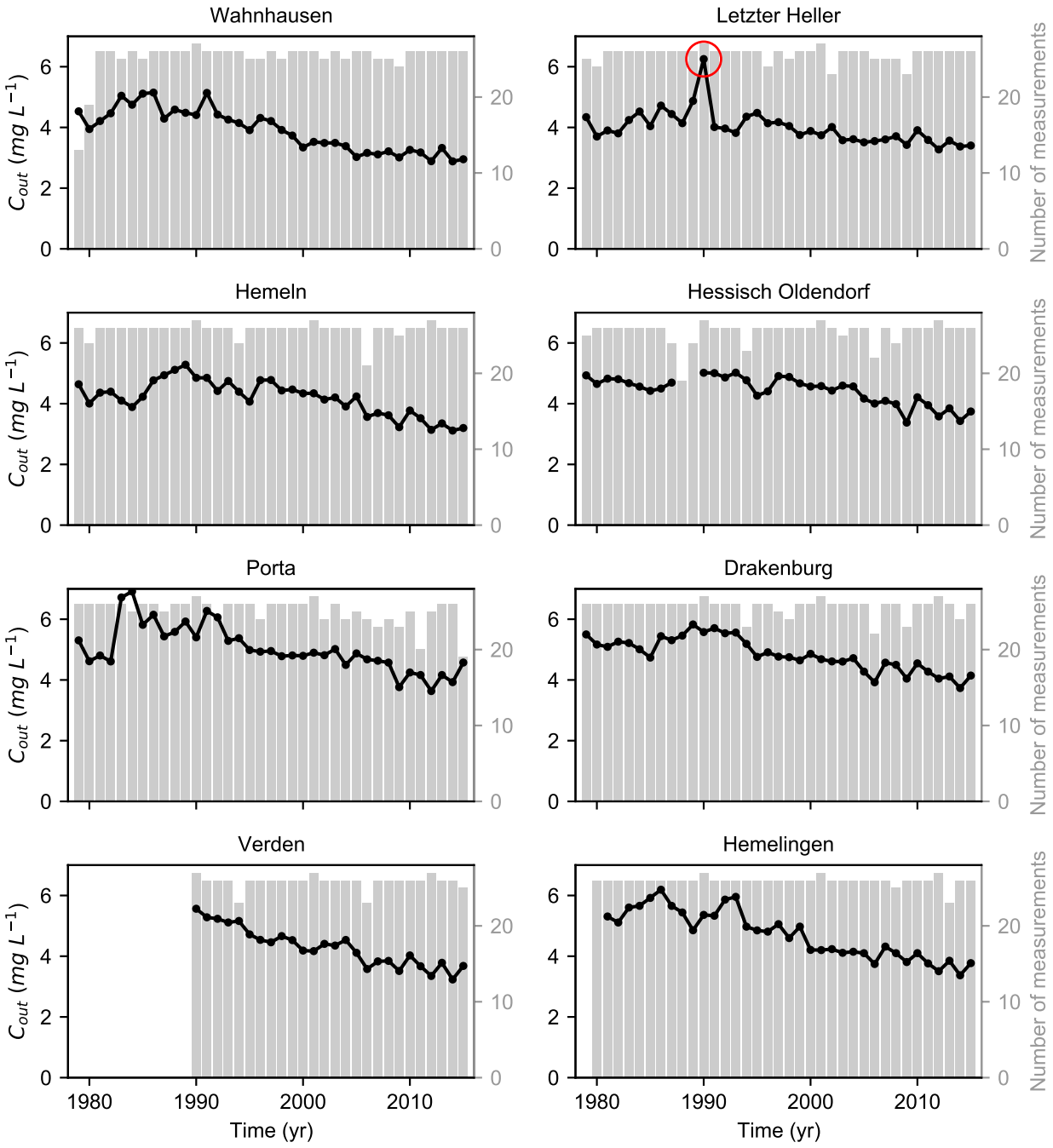

Figure S14. Time series of annual observed in-stream N-NO_3 concentration at the catchment outlet C_{out} and number of 14-day average measurements provided for each year by the River Basin Commission Weser (FGG Weser, 2021). For the Letzter Heller catchments, we combine the concentration measurements at the Letzter Heller station, that are available for the period 1979–2002, and at the Witzenhausen station, that are available for the period 2003–2015. The Witzenhausen station is located 8 km upstream of the Letzter Heller station. The red circle identifies the outlier value at the Letzter Heller station that we do not consider for comparison with ELEMeNT simulations.

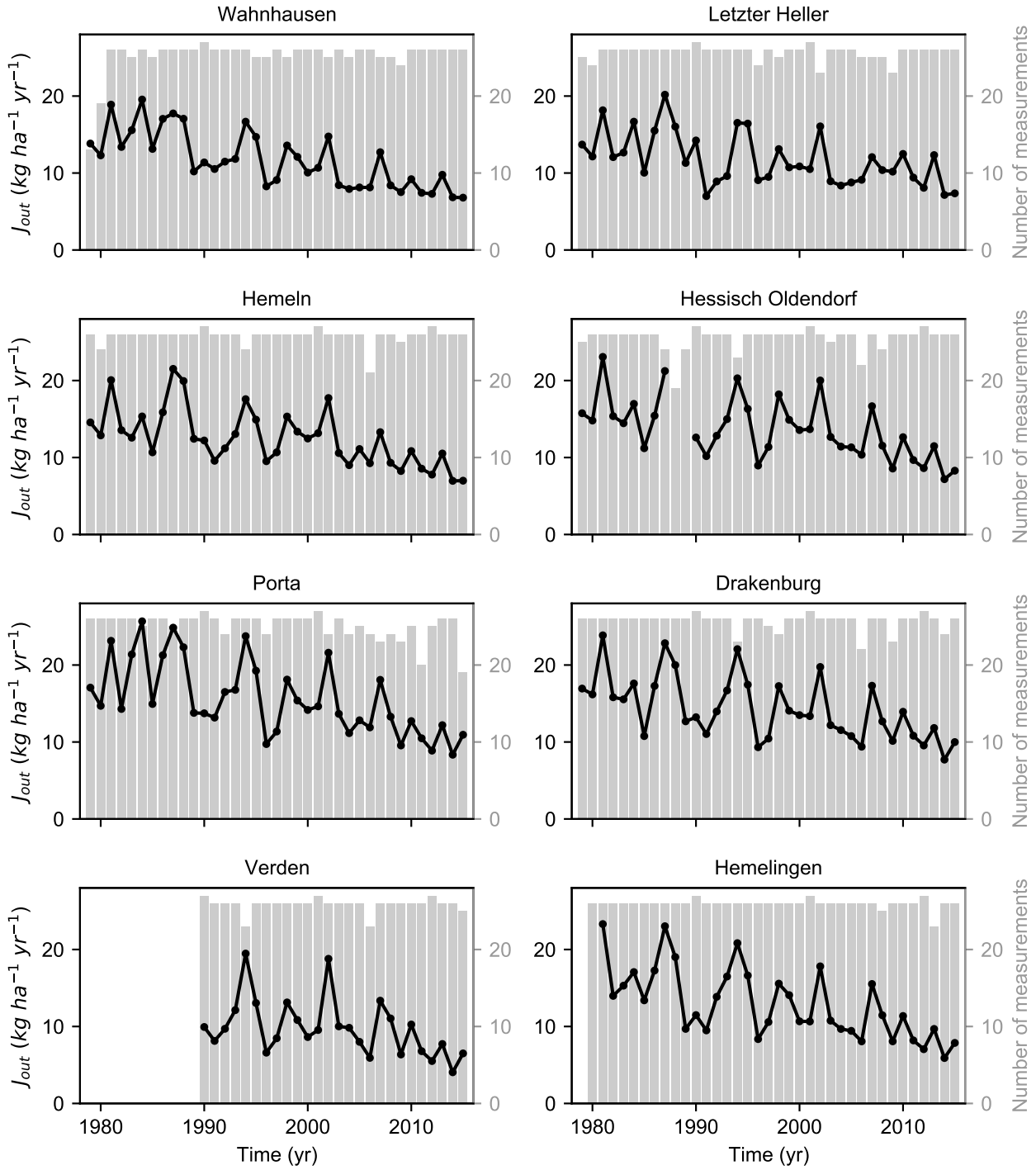

Figure S15. Time series of annual average observed in-stream N-NO_3 loading at the catchment outlet J_{out} and number of 14-day average measurements provided for each year by the River Basin Commission Weser (FGG Weser, 2021). For the Letzter Heller catchments, we combine the concentration measurements at the Letzter Heller station, that are available for the period 1979–2002, and at the Witzenhausen station, that are available for the period 2003–2015. The Witzenhausen station is located 8 km upstream of the Letzter Heller station.

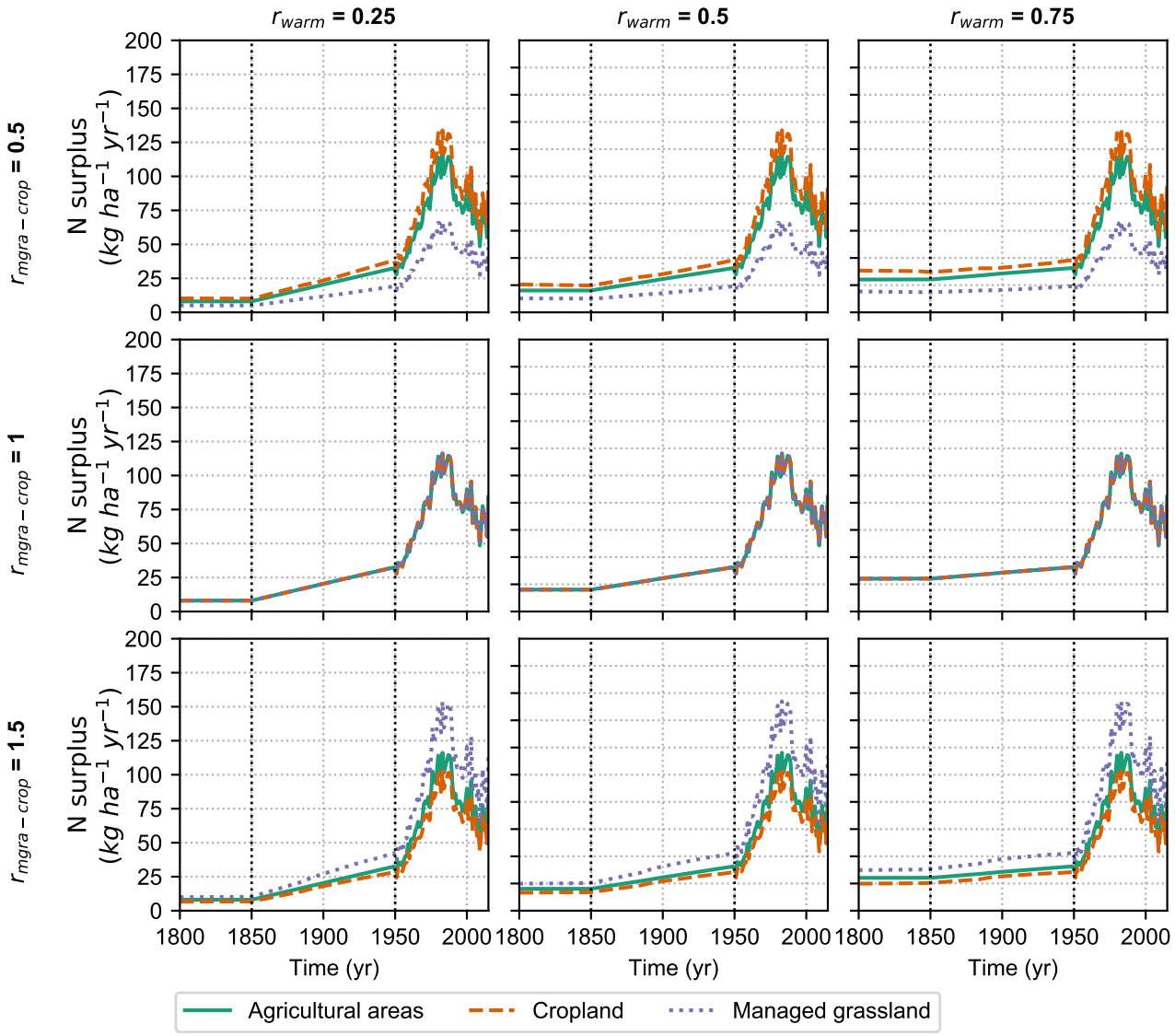

Figure S16. N surplus for total agricultural areas, cropland and agricultural permanent grassland for the period 1800–2015 for nine scenarios defined in this study for **Hemelingen**, corresponding to a value of the N surplus multiplier ($f_{surplus}$) equal to 1. The figure reports the corresponding values of the ratio of the N surplus for agricultural permanent grassland to the N surplus for cropland ($r_{mgra-crop}$), and the values of the ratio of the agricultural N surplus in 1850 to the value in 1950 (r_{warm}). Our "baseline" scenario is the central plot (i.e., $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$).

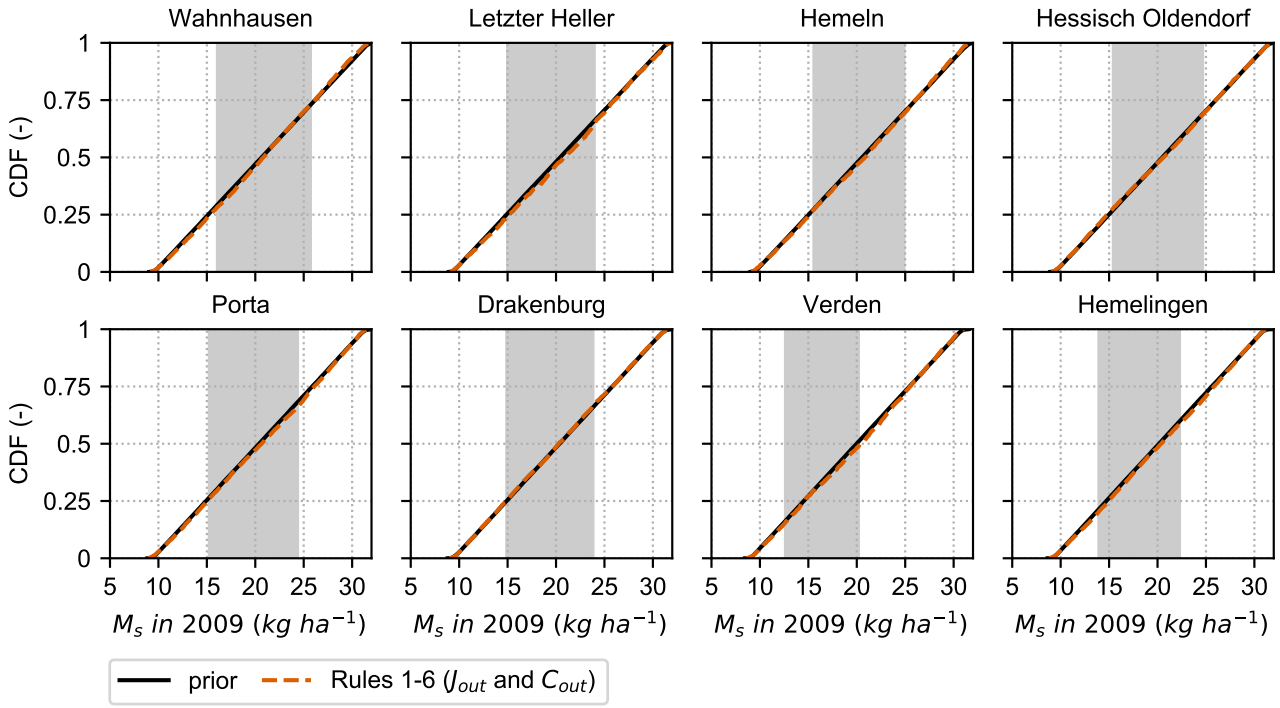

Figure S17. Cumulative Distribution Function (CDF) of the source zone N content M_s in 2009 in the initial simulation ensemble (100,000 realizations) and after application of rules 1-6 (rules on the in-stream loading J_{out} and concentration C_{out}) for the eight subcatchments. The grey shaded areas indicate the plausible ranges reported in Table S5, which are used in the definition of the rule on the source zone N content (rule 7).

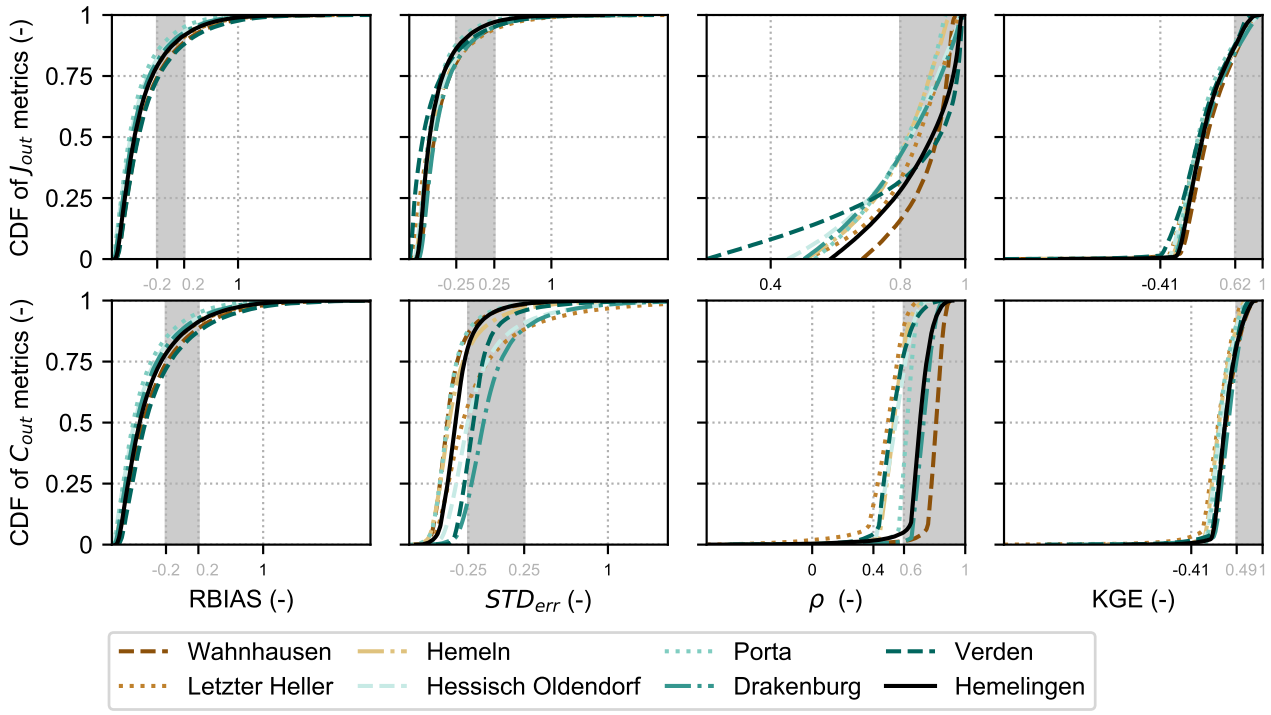

Figure S18. Application of the soft rules for the eight subcatchments: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream N loading (J_{out}) and concentration (C_{out}) in the initial simulation ensemble (100,000 realizations). The figure reports the three performance metrics used in the definition of rules 1-6 (the relative bias *RBIAS*, the variability error STD_{err} and the Pearson correlation coefficient ρ) and the Kling-Gupta efficiency (*KGE*). The grey shaded areas and grey numbers on the x-axis indicate the behavioural ranges of the performance metrics used in the definition of rules 1-6. The figure shows the **full ranges** of variation of the performance metrics in the simulation ensemble.

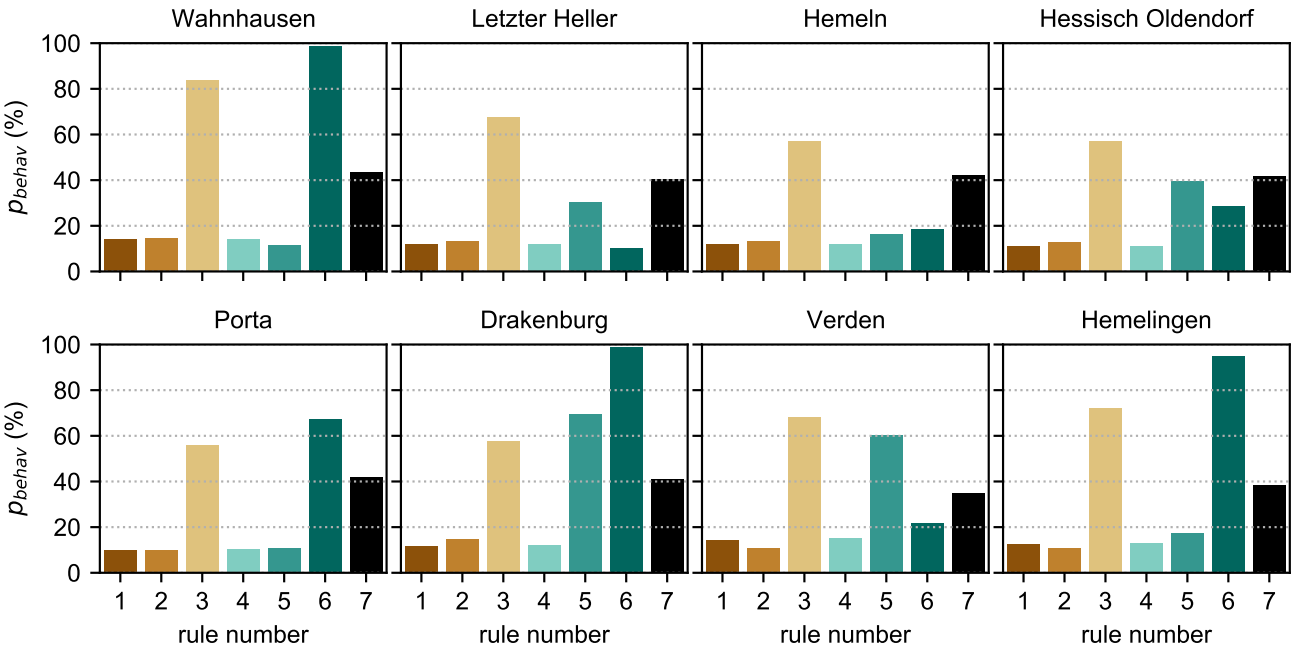

Figure S19. Percentage of behavioural model realizations when applying each of the seven rules individually for the eight subcatchments.

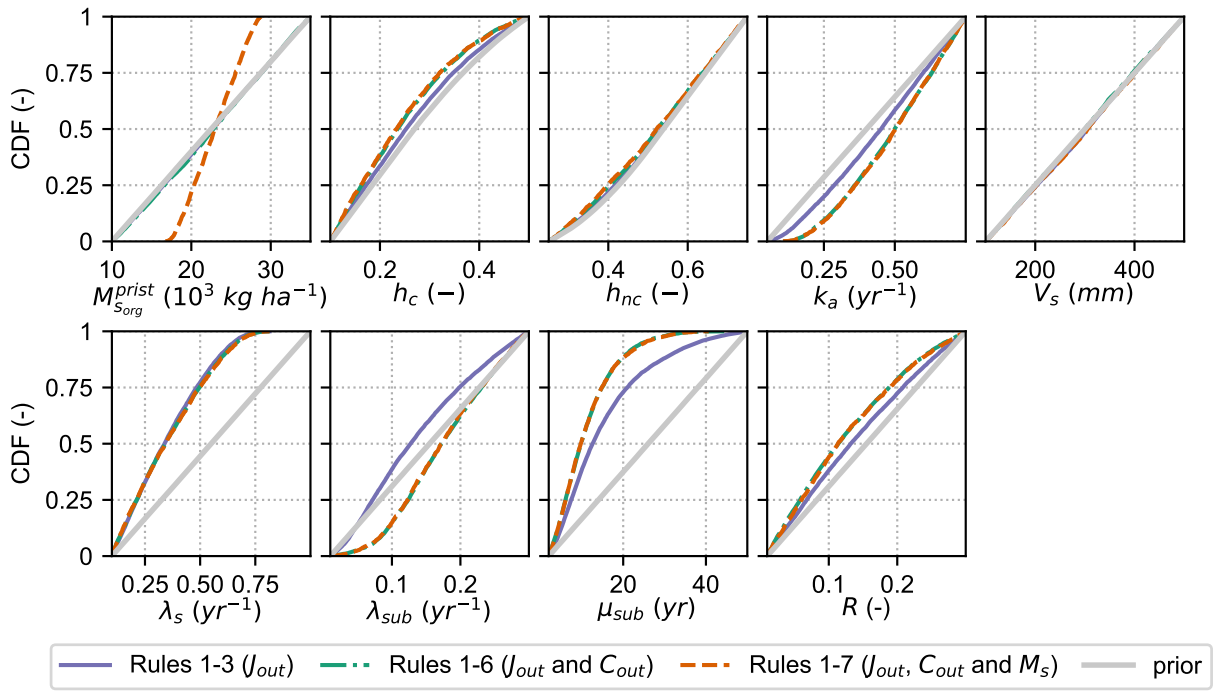

Figure S20. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Wahnhausen**.

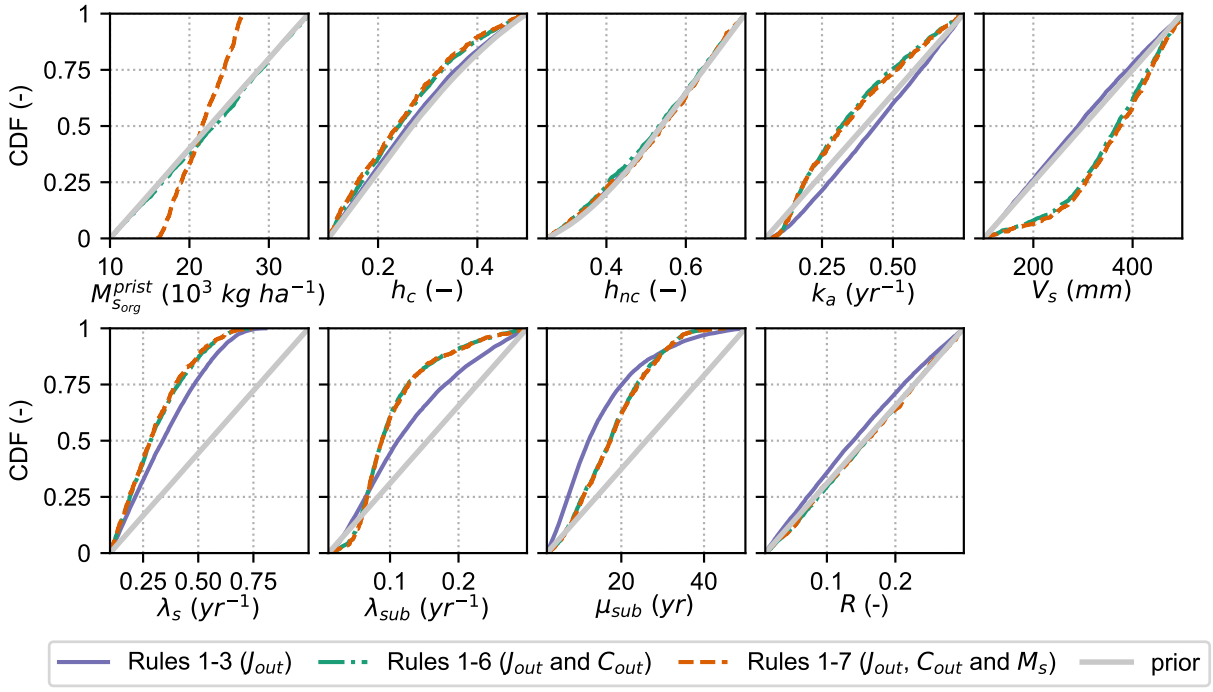

Figure S21. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Letzter Heller**.

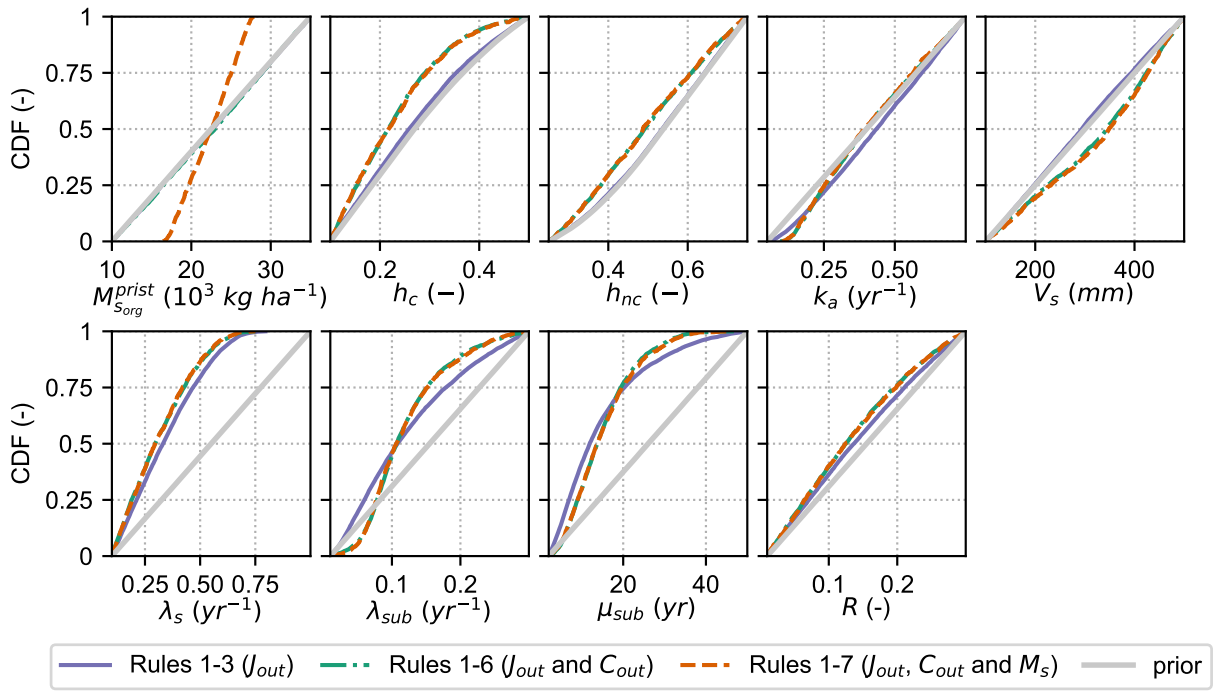

Figure S22. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hemeln**.

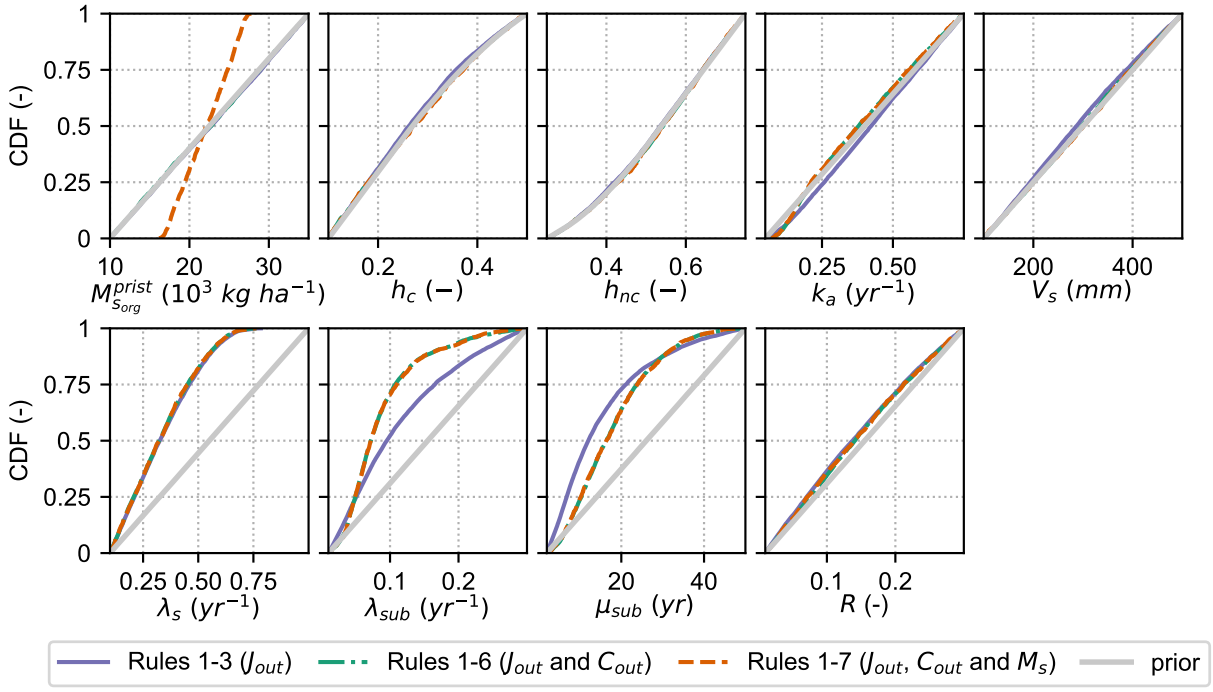

Figure S23. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hessisch Oldendorf**.

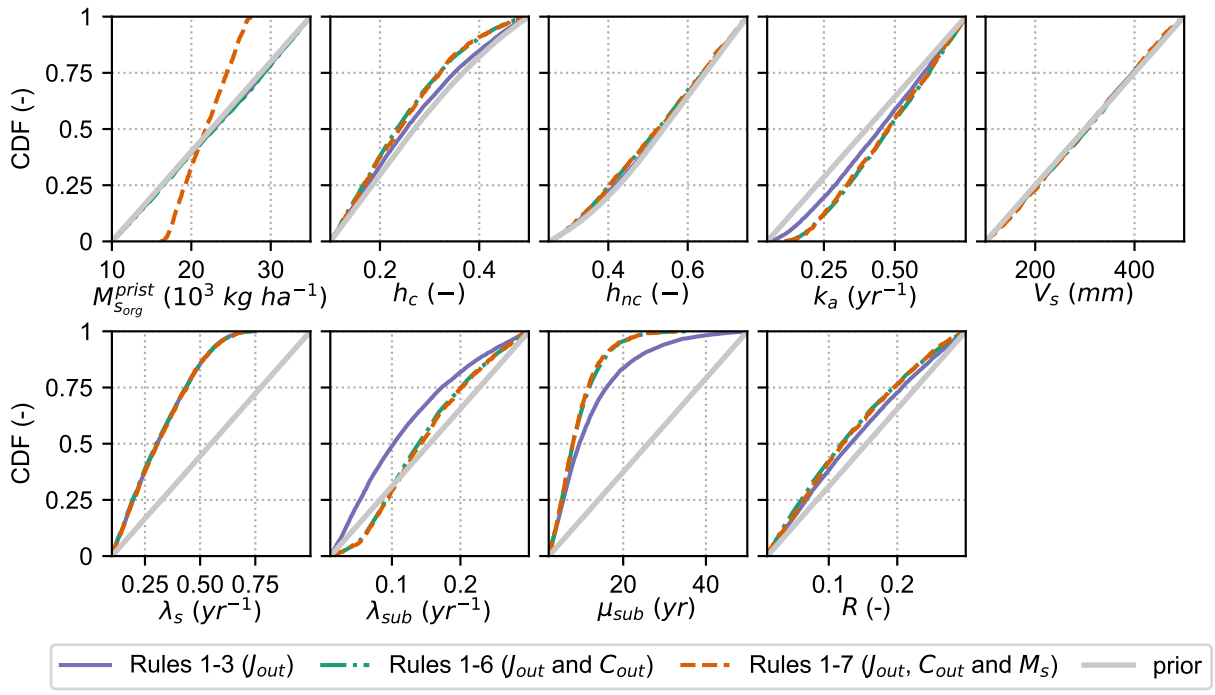

Figure S24. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Porta**.

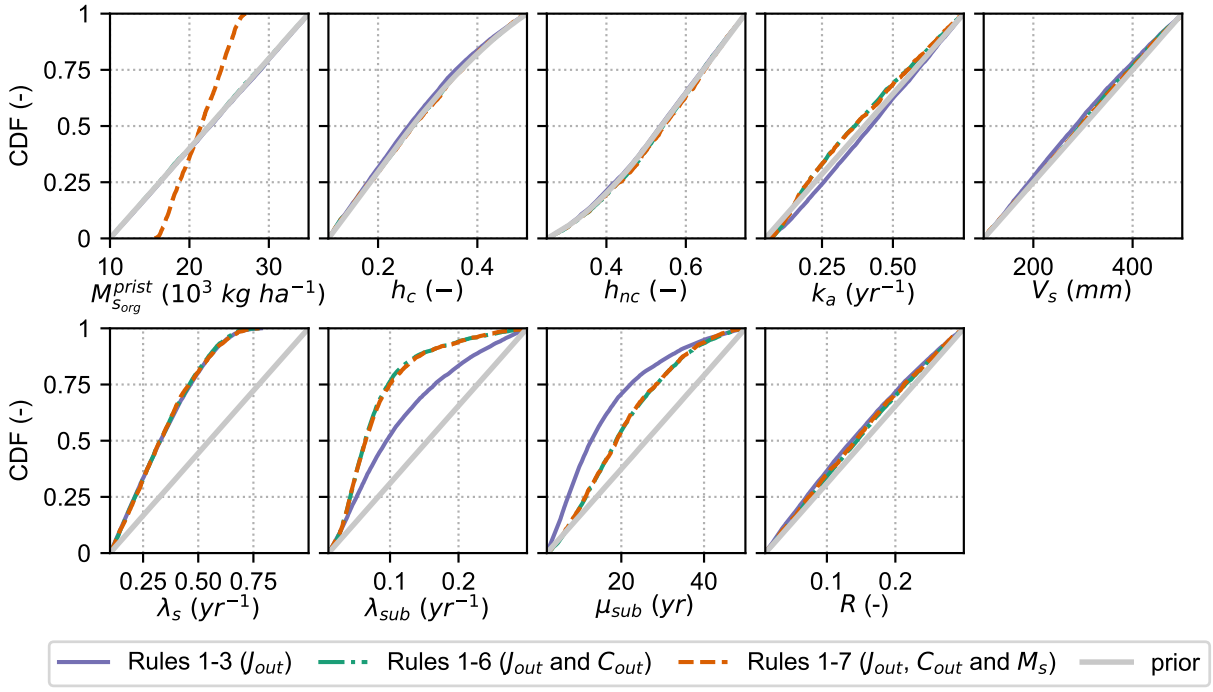

Figure S25. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Drakenburg**.

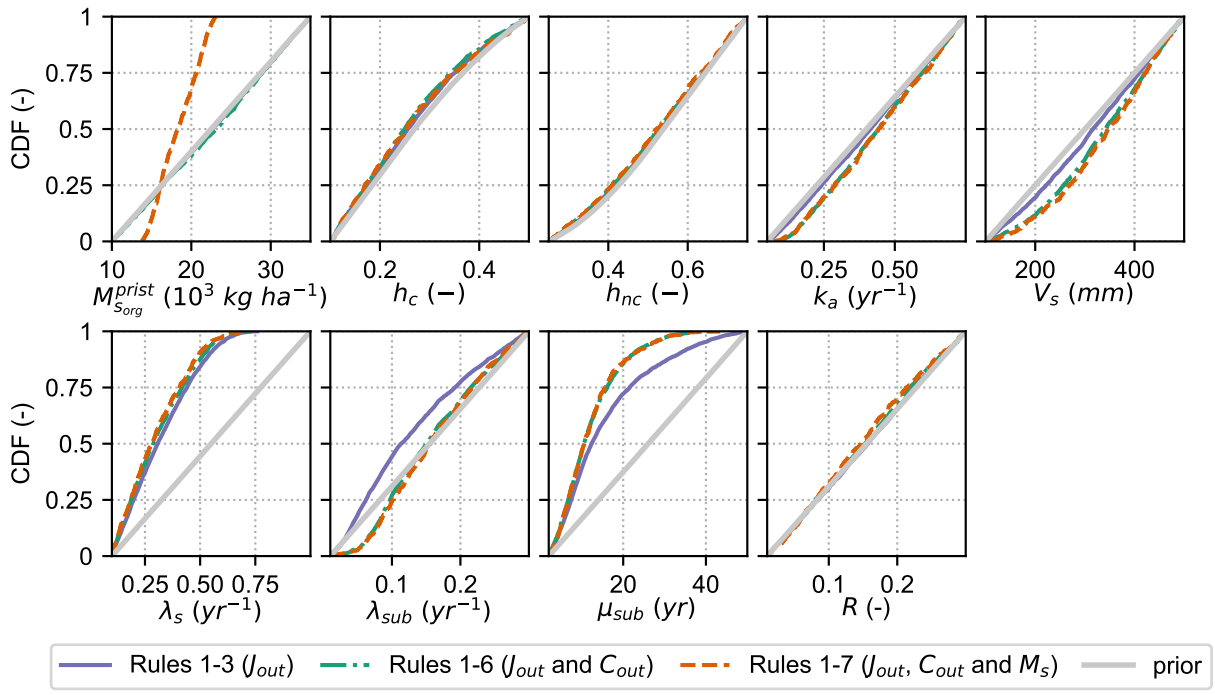

Figure S26. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Verden**.

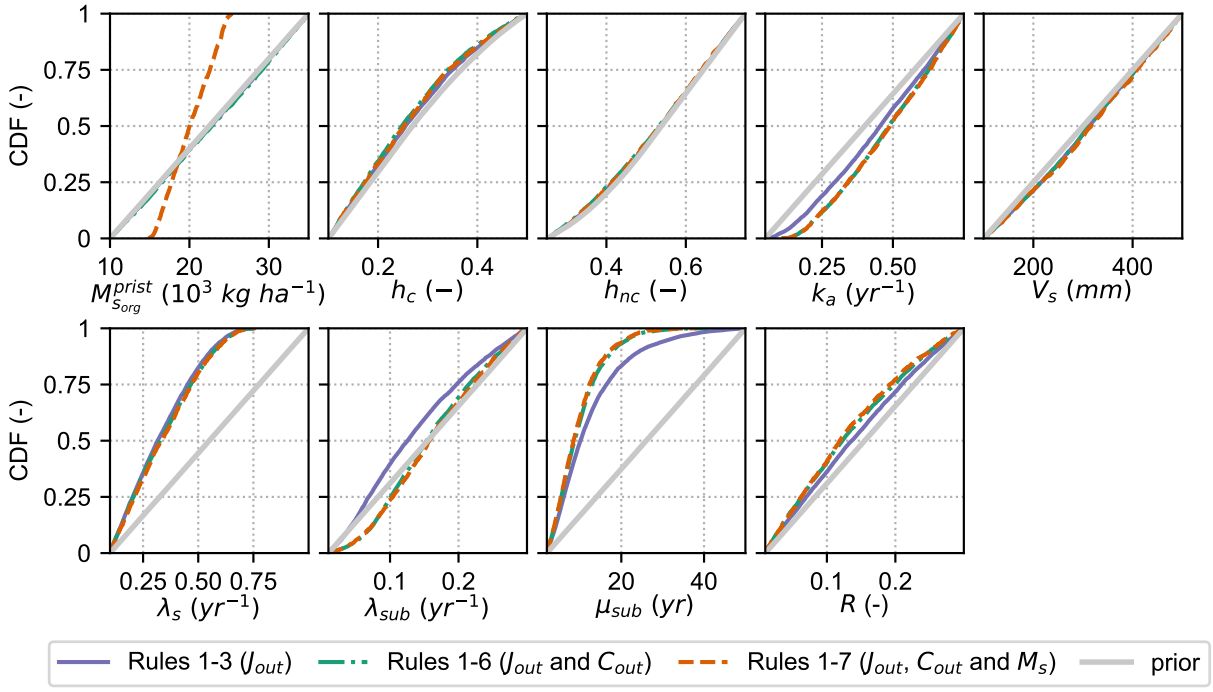

Figure S27. Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hemelingen**.

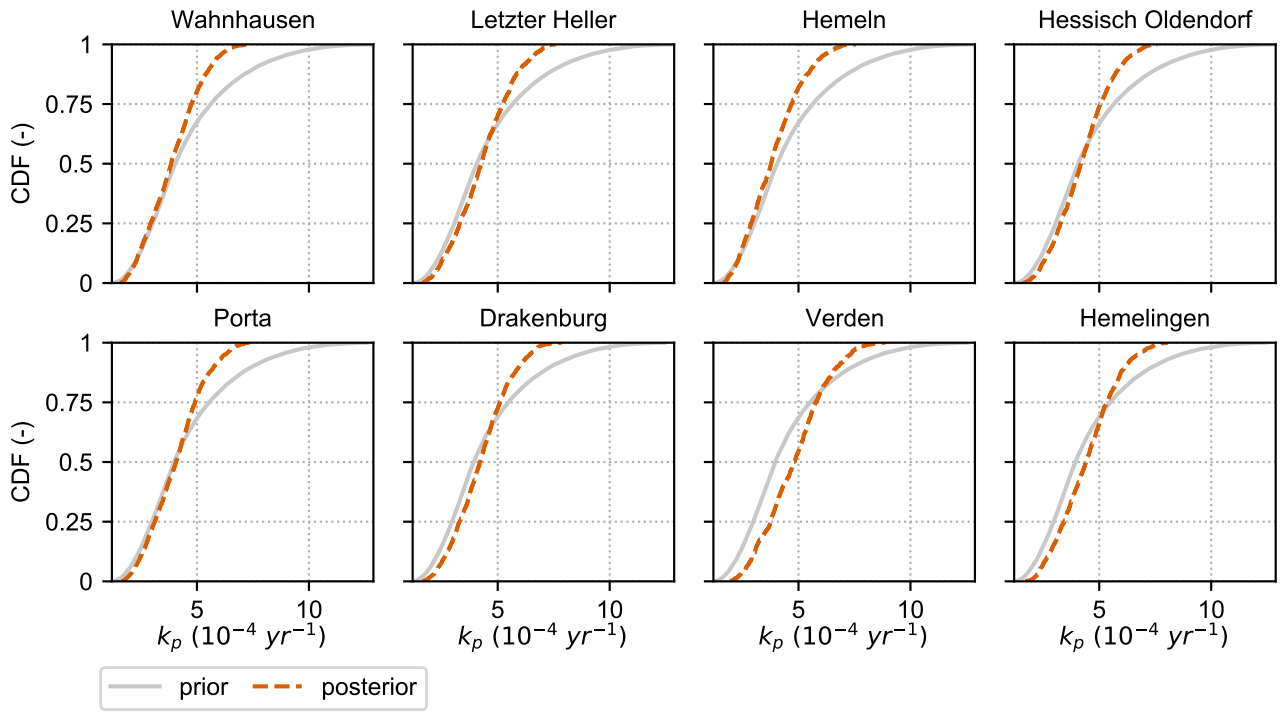

Figure S28. Cumulative Distribution Function (CDF) of the mineralization rate constant for the source zone organic protected N pool (k_p, calculated from Equation S24) in the prior simulation ensemble (before application of the soft rules) and in the posterior simulation ensemble (after application of the soft rules) for the eight subcatchments.

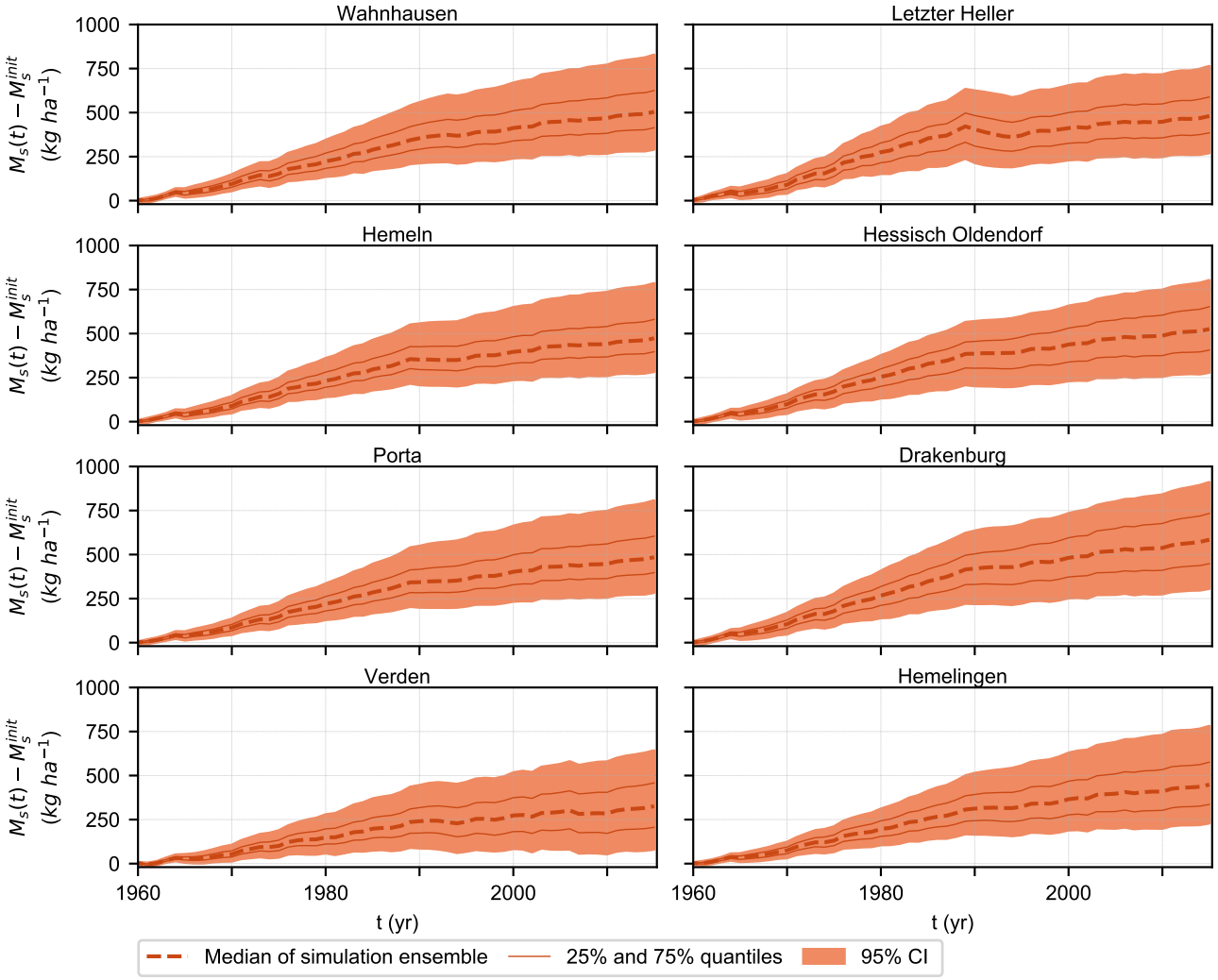

Figure S29. Simulated cumulative change in N storage in the source zone since 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation ensemble. Notations: t is the time; M_s is the source zone storage; M_s^{init} is the initial condition for the source zone storage in 1960.

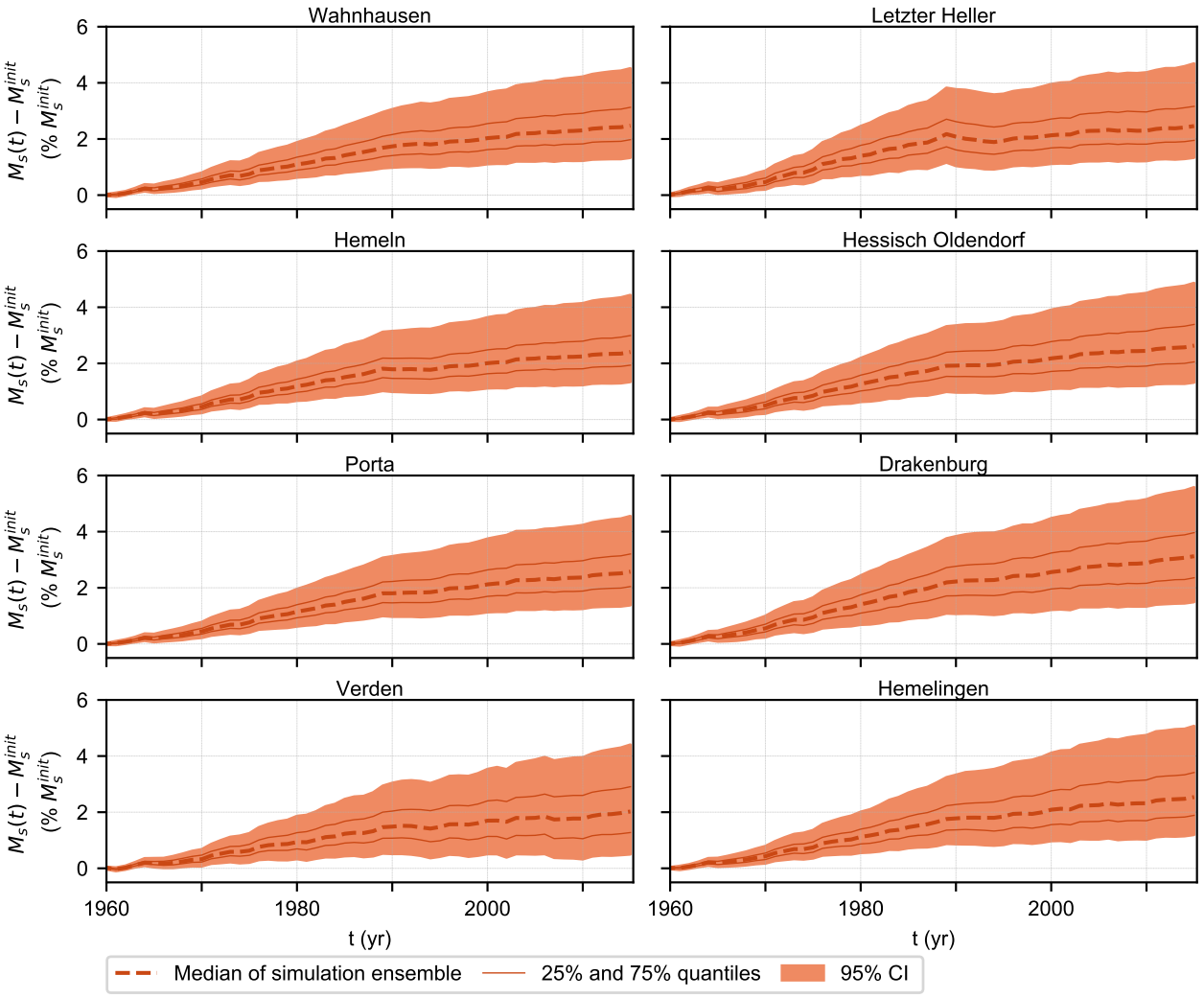

Figure S30. Simulated cumulative change in N storage in the source zone since 1960 as a percentage of the initial storage in 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation ensemble. Notations: t is the time; M_s is the source zone storage; M_s^{init} is the initial condition for the source zone storage in 1960.

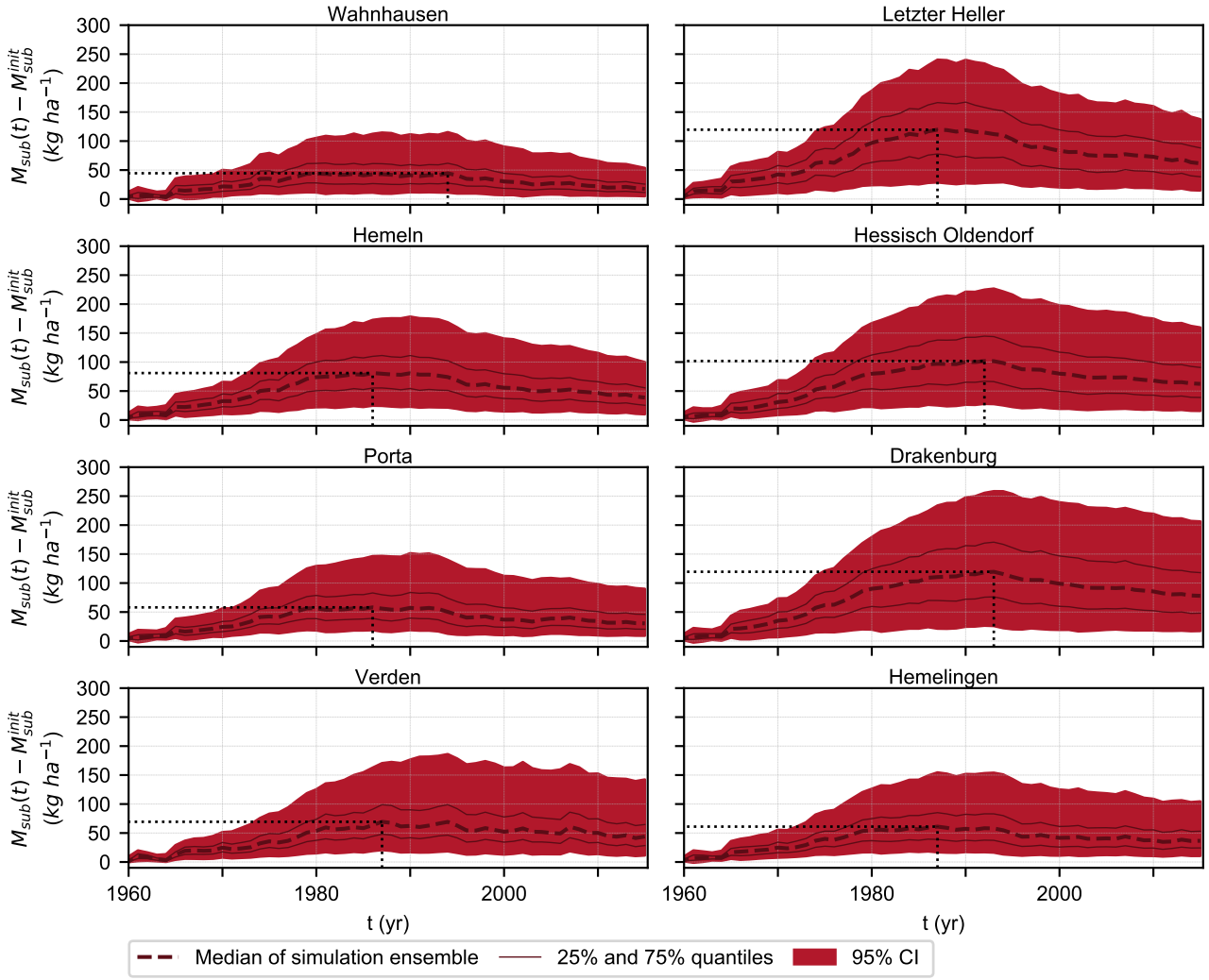

Figure S31. Simulated cumulative change in N storage in the subsurface since 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation ensemble. The maximum value of the median time series and corresponding year are indicated with black dotted lines. Notations: t is the time; M_{sub} is the subsurface zone storage; M_{sub}^{init} is the initial condition for the subsurface storage in 1960.

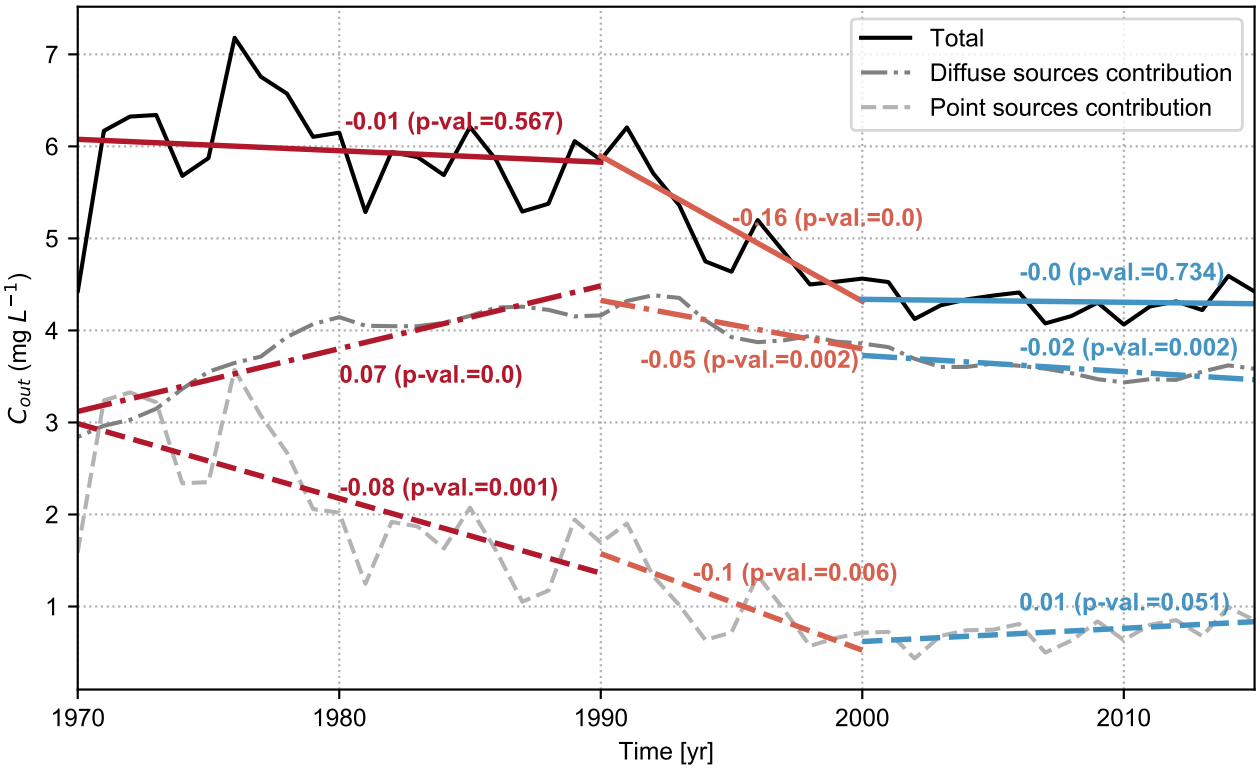

Figure S32. Time series of the simulated in-stream N concentration C_{out} (median value of the behavioural simulation ensemble, grey lines) and linear regression lines (coloured lines) for the period 1970–2015 over the Weser River Basin (at Hemelingen). The figure reports the total N concentration, and its two constituents, namely the contributions resulting from the N diffuse sources (N coming from the subsurface, $C_{out_{sub}}$) and from the N point sources ($C_{out_{ps}}$) ($C_{out} = C_{out_{sub}} + C_{out_{ps}}$). The regression lines were estimated for three different time periods to analyse the concentration trends (1970–1990, 1990–2000 and 2000–2015). The coloured numbers represent the slope for each regression line (in $mg\ L^{-1}\ yr^{-1}$) and the p-values of the linear trend are reported in brackets.

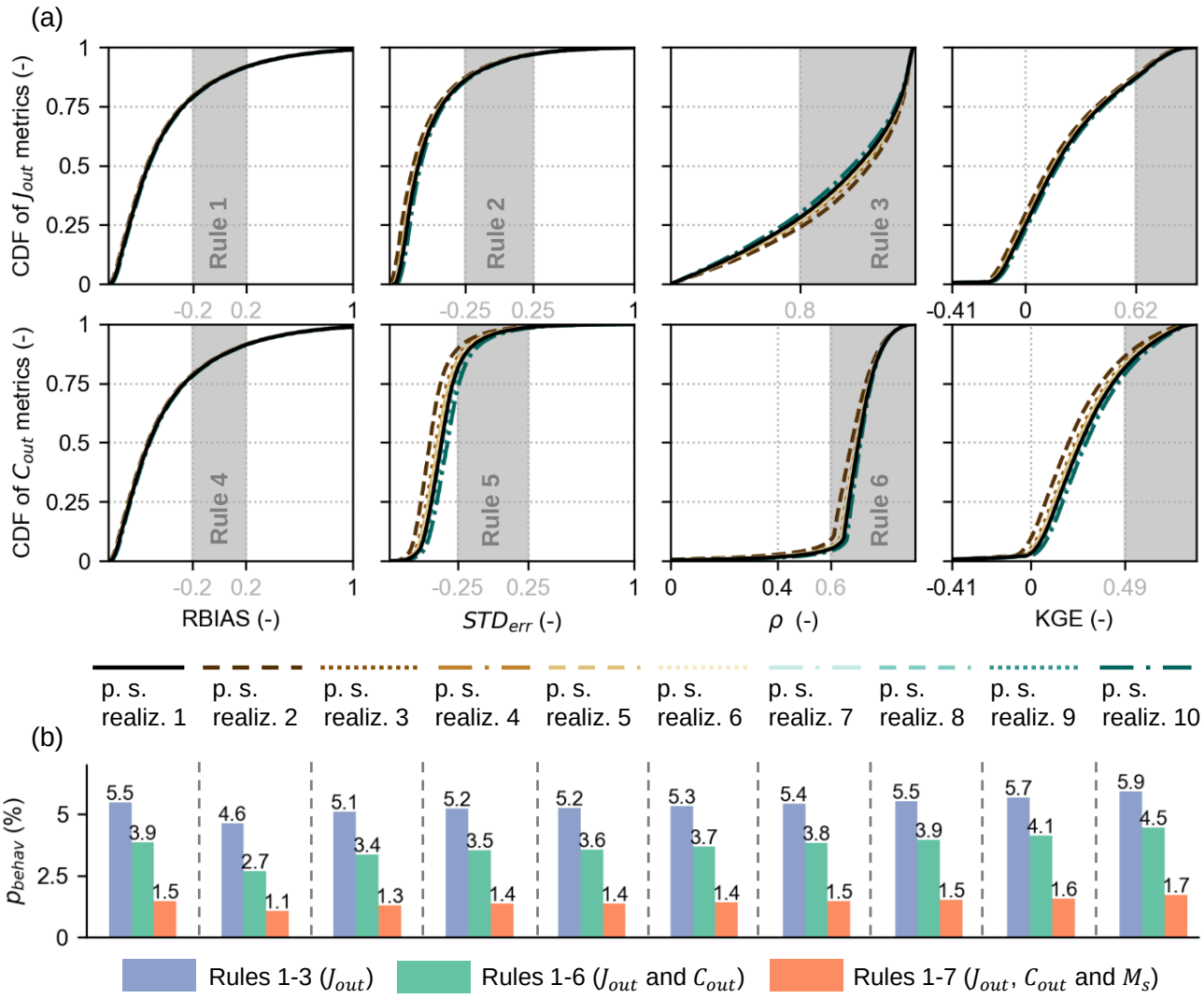

Figure S33. Application of the soft rules for the ten point source realizations (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading (J_{out}) and concentration (C_{out}) in the initial simulation ensemble (100,000 realizations) and (b) percentage of realizations of the initial ensemble identified as behavioural (p_{behav}) by successive application of the soft rules based on performance metrics for loading (J_{out}, rules 1-3) and concentration (C_{out}, rules 4-6), and based on the source zone N content for year 2009 (M_s, rule 7). Panel (a) reports the three performance metrics used in the definition of rules 1-6 (the relative bias $RBIAS$, the variability error STD_{err} and the Pearson correlation coefficient ρ) and the Kling-Gupta efficiency (KGE). The grey shaded areas and grey numbers on the x-axis indicate the behavioural ranges of the performance metrics used in the definition of rules 1-6. The range of the performance metrics shown do not include the extreme values.

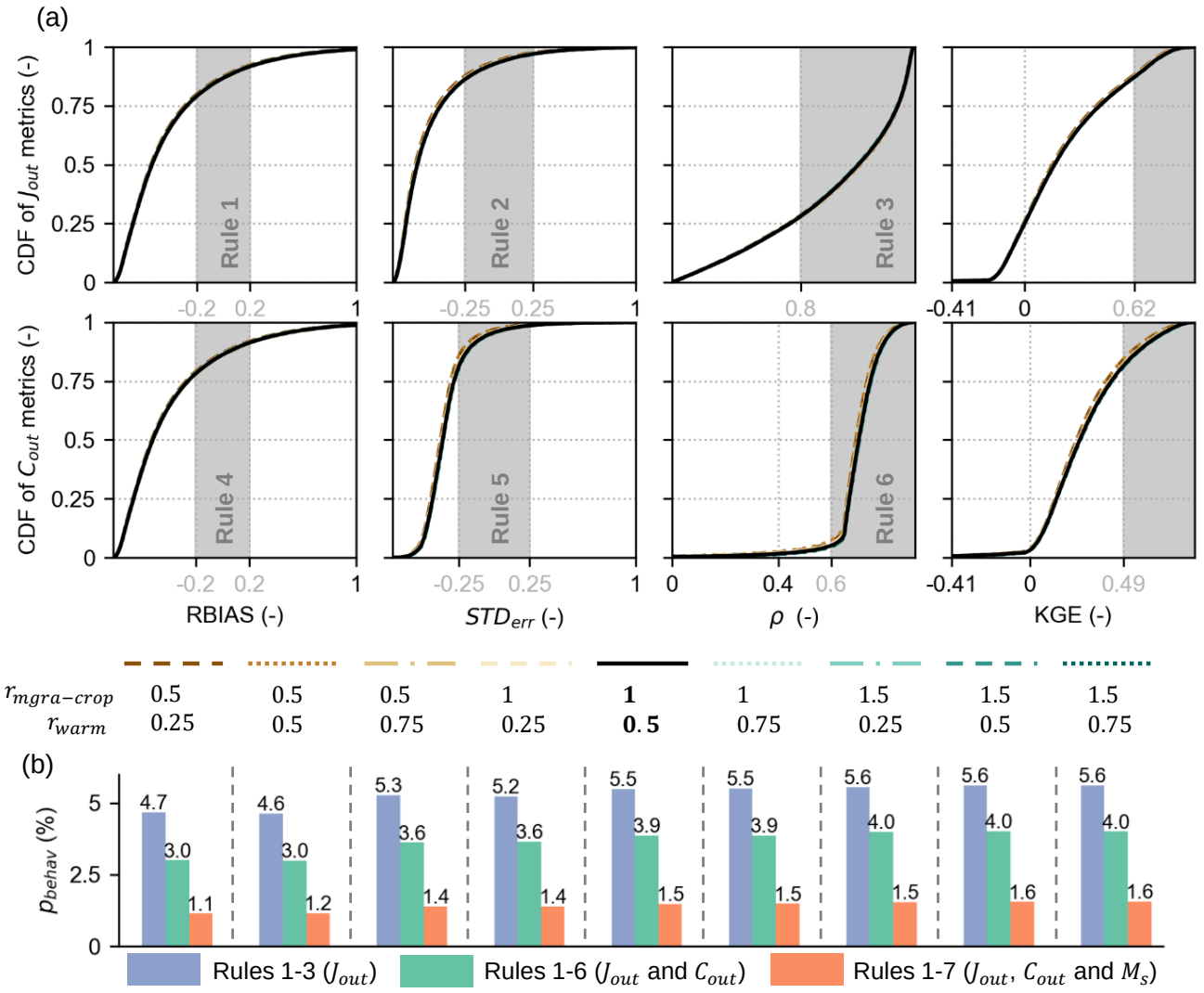

Figure S34. Application of the soft rules for the nine combinations of $r_{mgra-crop}$ and r_{warm} (for $f_{surplus} = 1$ and the first point source realization) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading (J_{out}) and concentration (C_{out}) in the initial simulation ensemble (100,000 realizations) and (b) percentage of realizations of the initial ensemble identified as behavioural (p_{behav}) by successive application of the soft rules based on performance metrics for loading (J_{out}, rules 1-3) and concentration (C_{out}, rules 4-6), and based on the source zone N content for year 2009 (M_s, rule 7). Panel (a) reports the three performance metrics used in the definition of rules 1-6 (the relative bias *RBIAS*, the variability error STD_{err} and the Pearson correlation coefficient ρ) and the Kling-Gupta efficiency (*KGE*). The grey shaded areas and grey numbers on the x-axis indicate the behavioural ranges of the performance metrics used in the definition of rules 1-6. The range of the performance metrics shown do not include the extreme values.

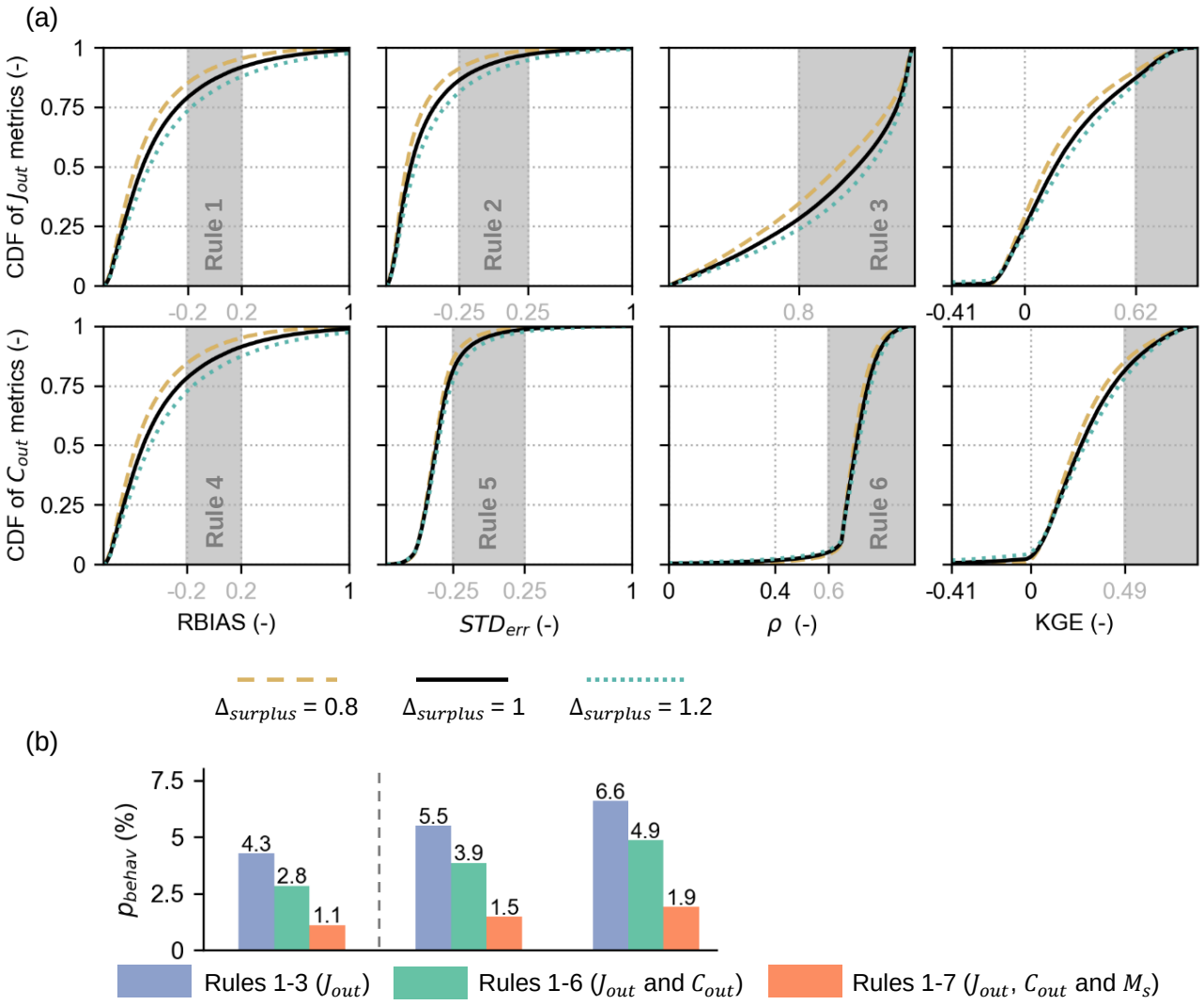

Figure S35. Application of the soft rules for the three realizations of $f_{surplus}$ (for $r_{mgra-crop} = 1$, $r_{warm} = 0.5$ and the first point source realization) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading (J_{out}) and concentration (C_{out}) in the initial simulation ensemble (100,000 realizations) and (b) percentage of realizations of the initial ensemble identified as behavioural (p_{behav}) by successive application of the soft rules based on performance metrics for loading (J_{out}, rules 1-3) and concentration (C_{out}, rules 4-6), and based on the source zone N content for year 2009 (M_s, rule 7). Panel (a) reports the three performance metrics used in the definition of rules 1-6 (the relative bias *RBIAS*, the variability error STD_{err} and the Pearson correlation coefficient ρ) and the Kling-Gupta efficiency (*KGE*). The grey shaded areas and grey numbers on the x-axis indicate the behavioural ranges of the performance metrics used in the definition of rules 1-6. The range of the performance metrics shown do not include the extreme values.

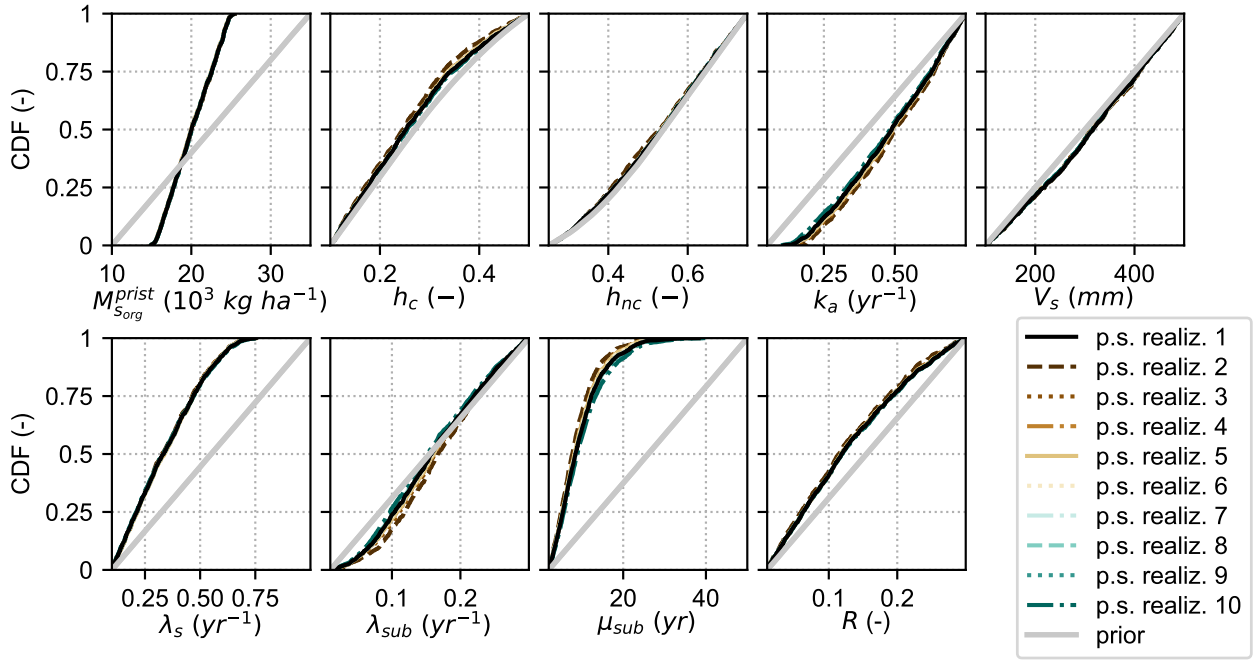

Figure S36. Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the ten N point sources realizations (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$), and prior CDF in the original parameter sample of size 100,000 (grey line), which is the same for all sets of simulations, for **Hemelingen**. The baseline scenario (i.e., first point source realization) is reported with a black solid line.

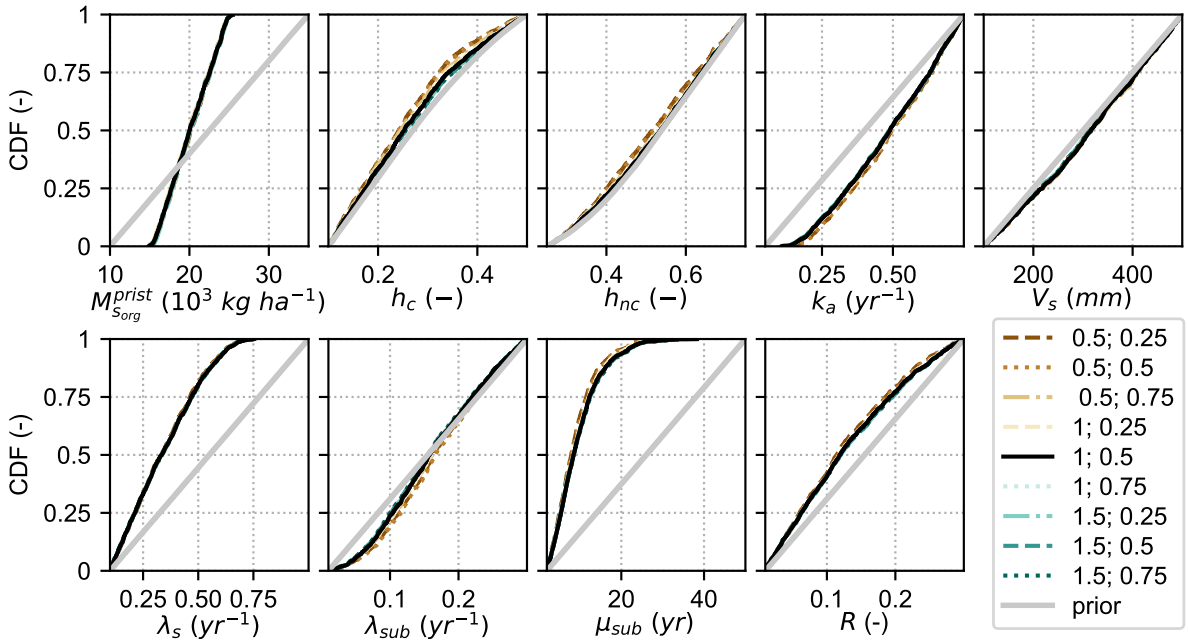

Figure S37. Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the nine combinations of $r_{mgra-crop}$ and r_{warm} (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, and the first point source realization), and prior CDF in the original parameter sample of size 100,000 (grey line), which is the same for all scenarios, for **Hemelingen**. The numbers in the legend indicate the value of the ratio of N surplus of agricultural permanent grassland to N surplus of cropland ($r_{mgra-crop}$) and the ratio of the value of the agricultural N surplus in 1850 to the value in 1950 (r_{warm}) respectively for the nine N surplus realizations. The baseline scenario (i.e., $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$) is reported with a black solid line.

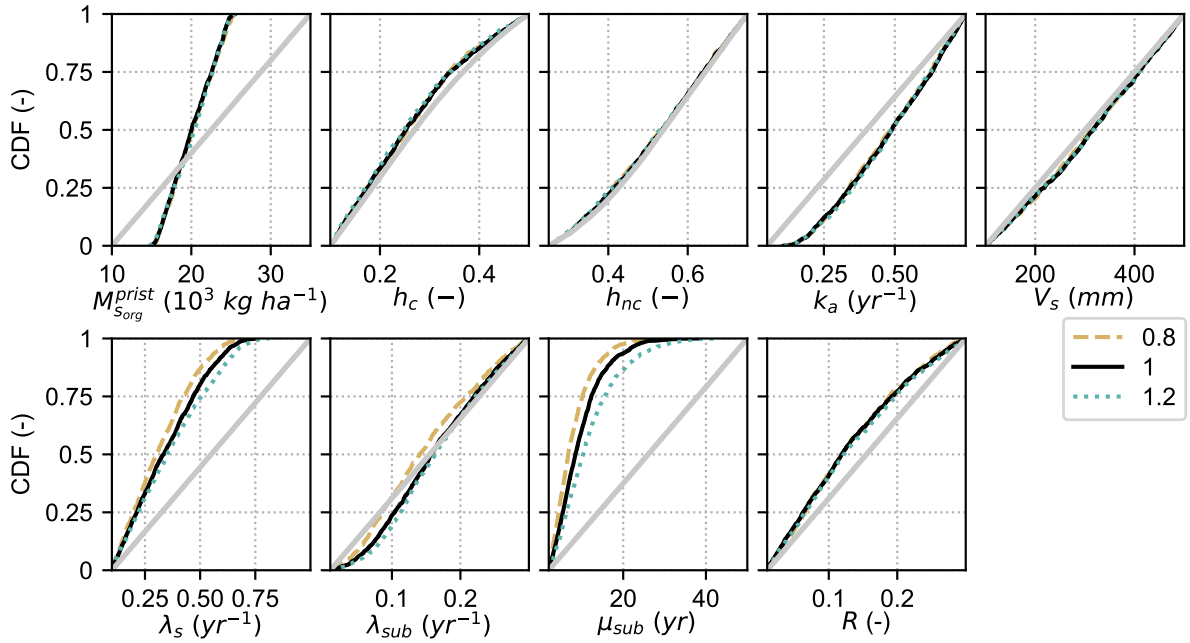

Figure S38. Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the three realizations of $f_{surplus}$ (for $r_{mgra-crop} = 1$, r_{warm}=0.5, and the first point source realization), and prior CDF in the original parameter sample of size 100,000 (grey line), which is the same for all scenarios, for **Hemelingen**. The numbers in the legend indicate the value of the ratio of N surplus of agricultural permanent grassland to N surplus of cropland ($r_{mgra-crop}$) and the ratio of the value of the agricultural N surplus in 1850 to the value in 1950 (r_{warm}) respectively for the nine N surplus realizations. The baseline scenario (i.e., $f_{surplus} = 1$) is reported with a black solid line.

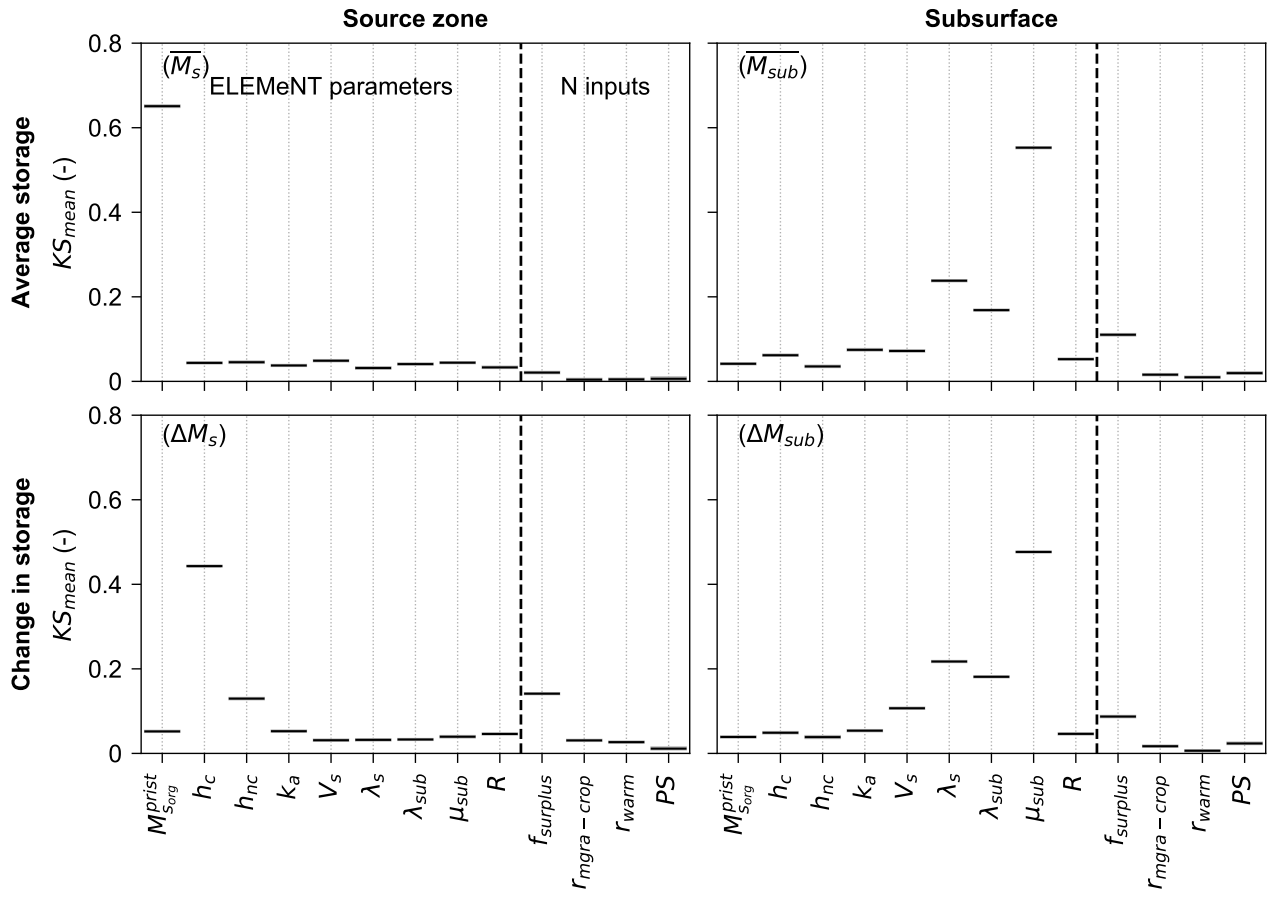

Figure S39. PAWN sensitivity indices (KS_{mean}, estimated as the mean value of KS statistic across all conditioning intervals) of the nine ELEMeNT parameters, the three parameters introduced to generate alternative N surplus realizations ($f_{surplus}$, $r_{mgra-crop}$, and r_{warm}), and the N point sources realization (PS), for the WRB at Hemelingen. Sensitivity indices are reported with respect to four model outputs evaluated over the period 1960–2015, namely the average source zone N storage $\overline{M_s}$, the average subsurface N storage $\overline{M_{sub}}$, the cumulative change in source zone N storage ΔM_s, and the cumulative change in subsurface N storage ΔM_{sub}. The horizontal black lines indicate the bootstrap mean value of the sensitivity indices, while the grey boxes represent the 95% bootstrap confidence intervals. The bootstrap confidence intervals are very small (the grey boxes are very narrow), since the size of the sample used to calculate the PAWN indices is very large.

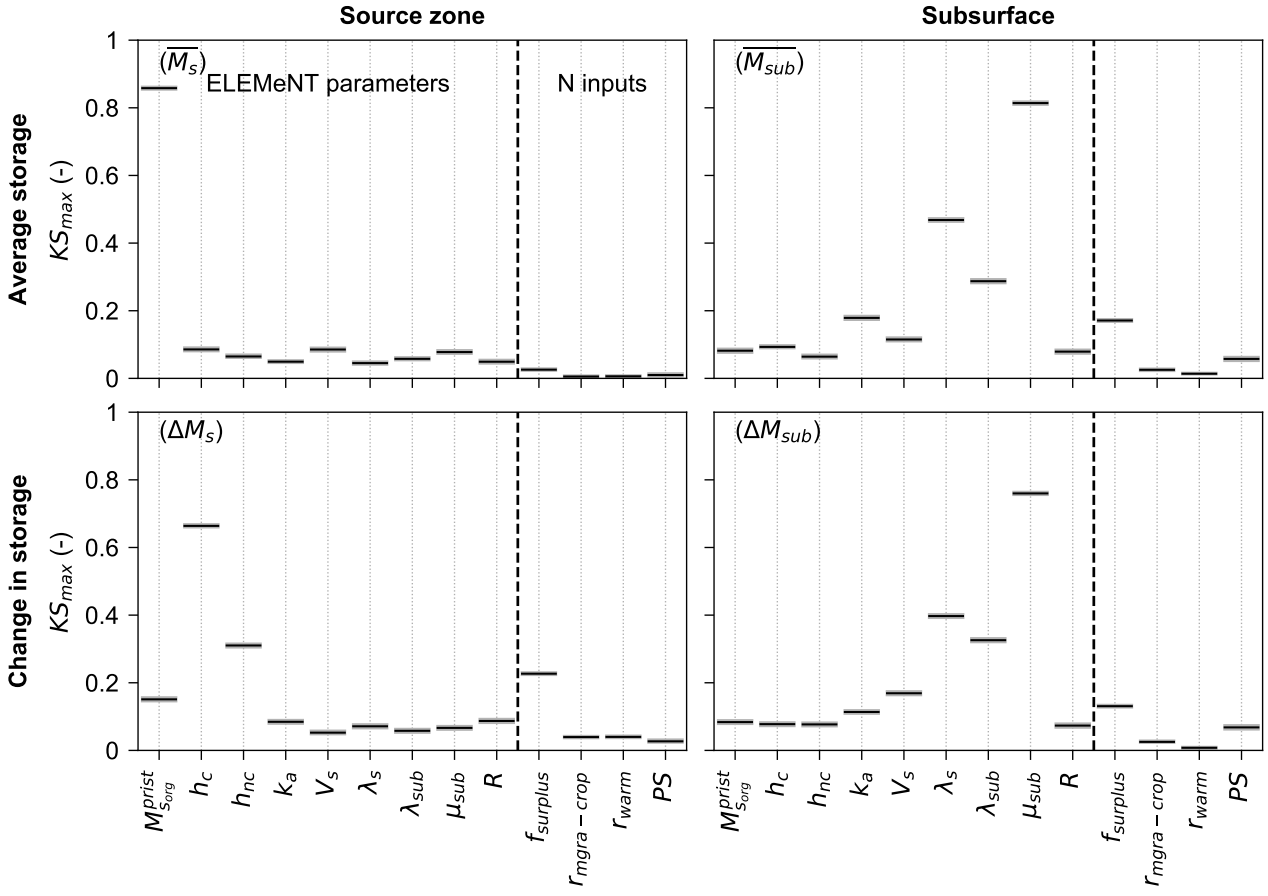

Figure S40. PAWN sensitivity indices (KS_{max}, estimated as the maximum value of KS statistic across all conditioning intervals) of the nine ELEMeNT parameters, the three parameters introduced to generate alternative N surplus realizations ($f_{surplus}$, $r_{mgra-crop}$, and r_{warm}), and the N point sources realization (PS), for the WRB at Hemelingen. Sensitivity indices are reported with respect to four model outputs evaluated over the period 1960–2015, namely the average source zone N storage $\overline{M_s}$, the average subsurface N storage $\overline{M_{sub}}$, the cumulative change in source zone N storage ΔM_s, and the cumulative change in subsurface N storage ΔM_{sub}. The horizontal black lines indicate the bootstrap mean value of the sensitivity indices, while the grey boxes represent the 95% bootstrap confidence intervals. The bootstrap confidence intervals are very small (the grey boxes are very narrow), since the size of the sample used to calculate the PAWN indices is very large.

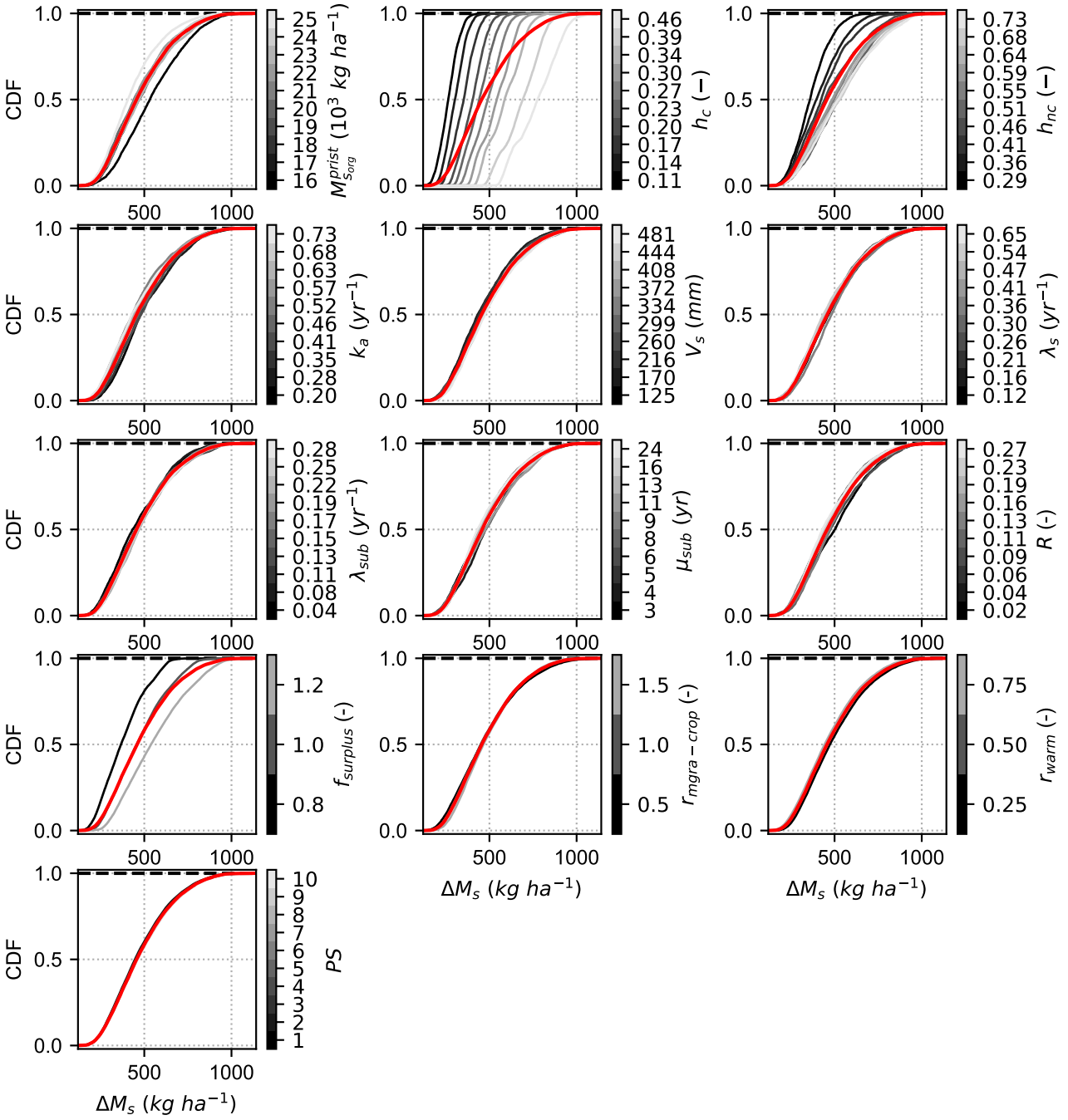

Figure S41. Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the change in source zone N storage over the period 1960–2015 (ΔM_s), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios and the N point sources realization (PS). These CDFs are used in the PAWN sensitivity analysis method. The colorbars report the average value of the parameters over each conditioning interval.

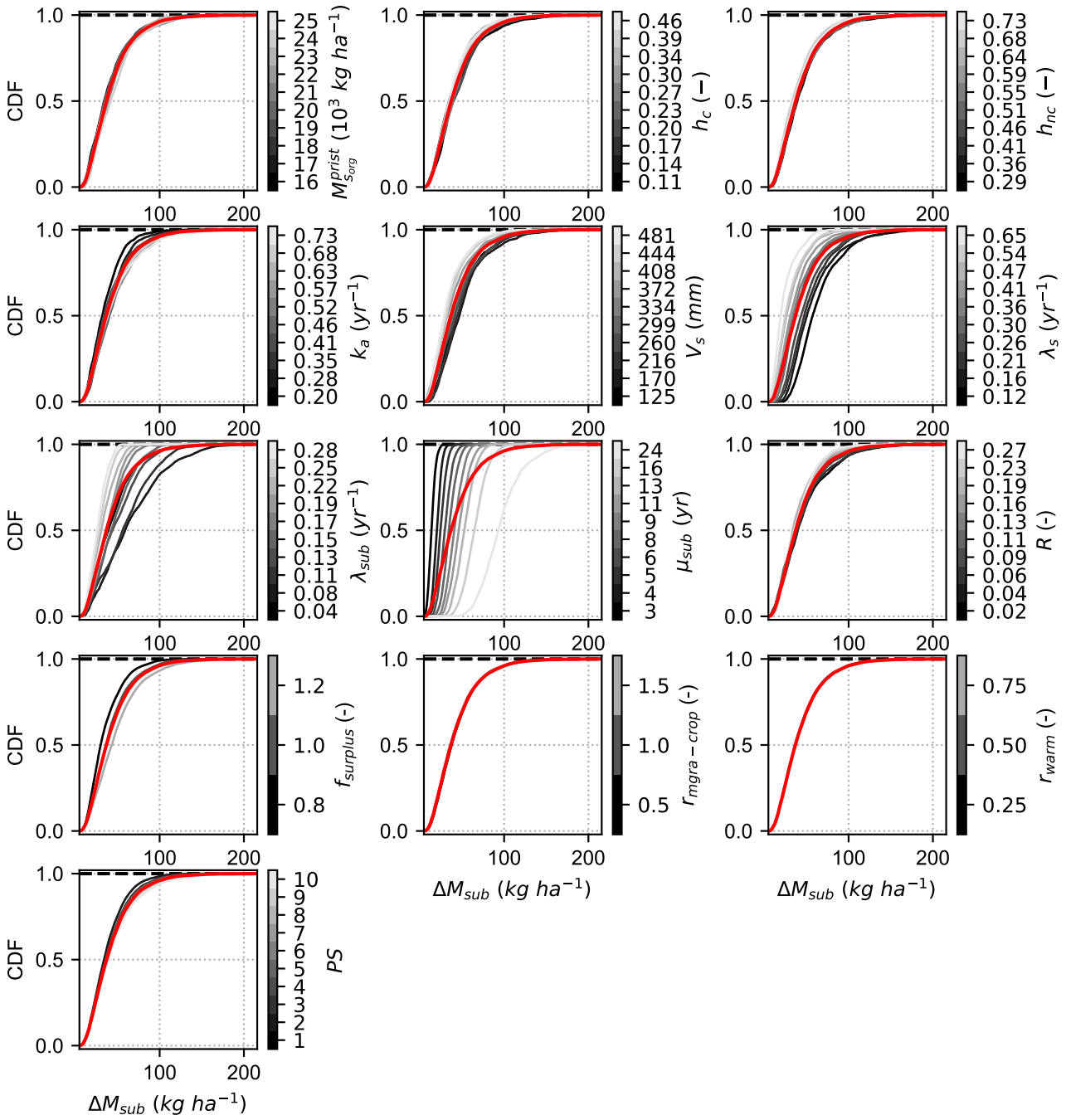

Figure S42. Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the change in subsurface N storage over the period 1960–2015 (ΔM_{sub}), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios and the N point sources realization (PS). These CDFs are used in the PAWN sensitivity analysis method. The colorbars report the average value of the parameters over each conditioning interval.

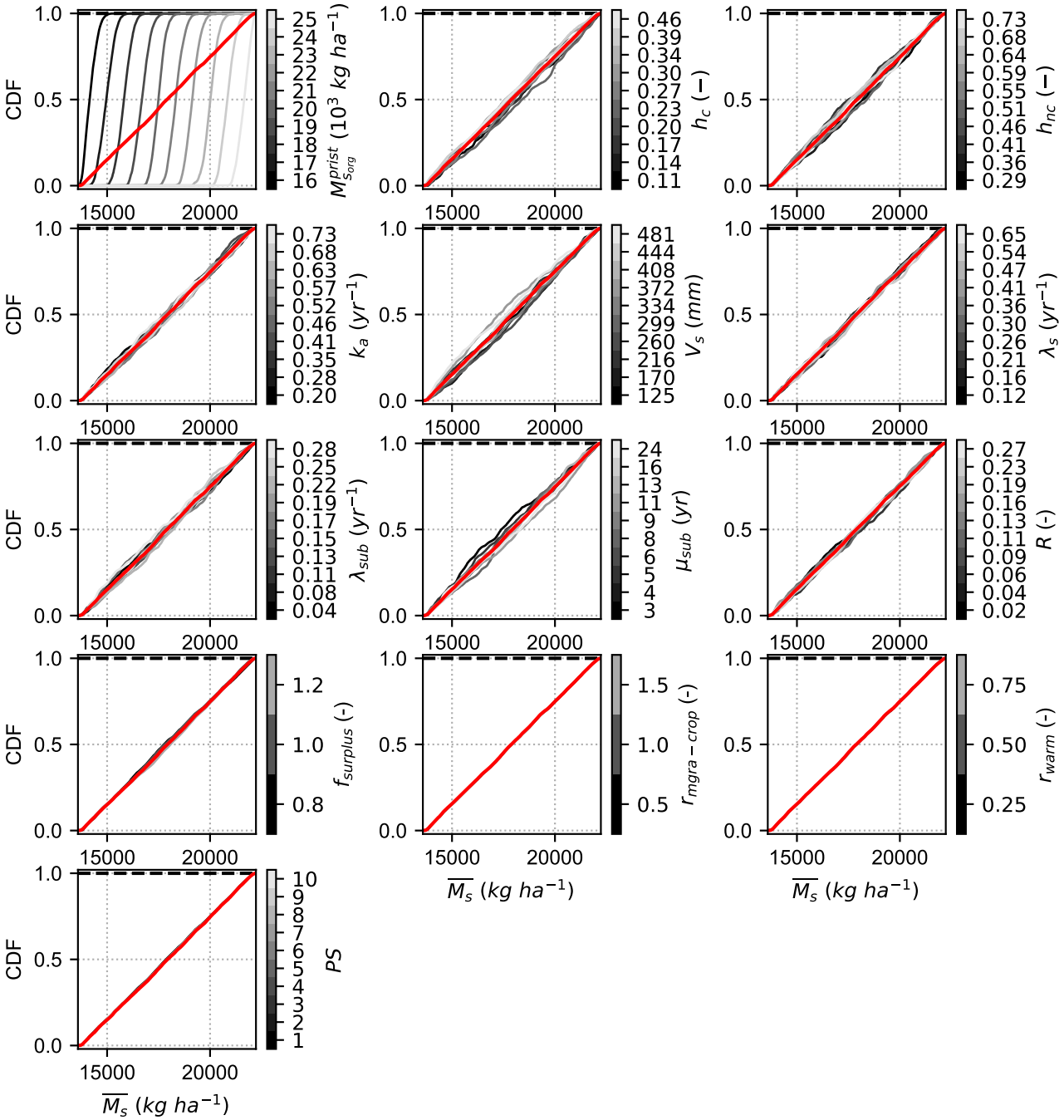

Figure S43. Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the average source zone N storage over the period 1960–2015 ($\overline{M_s}$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios and the N point sources realization (PS). These CDFs are used in the PAWN sensitivity analysis method. The colorbars report the average value of the parameters over each conditioning interval.

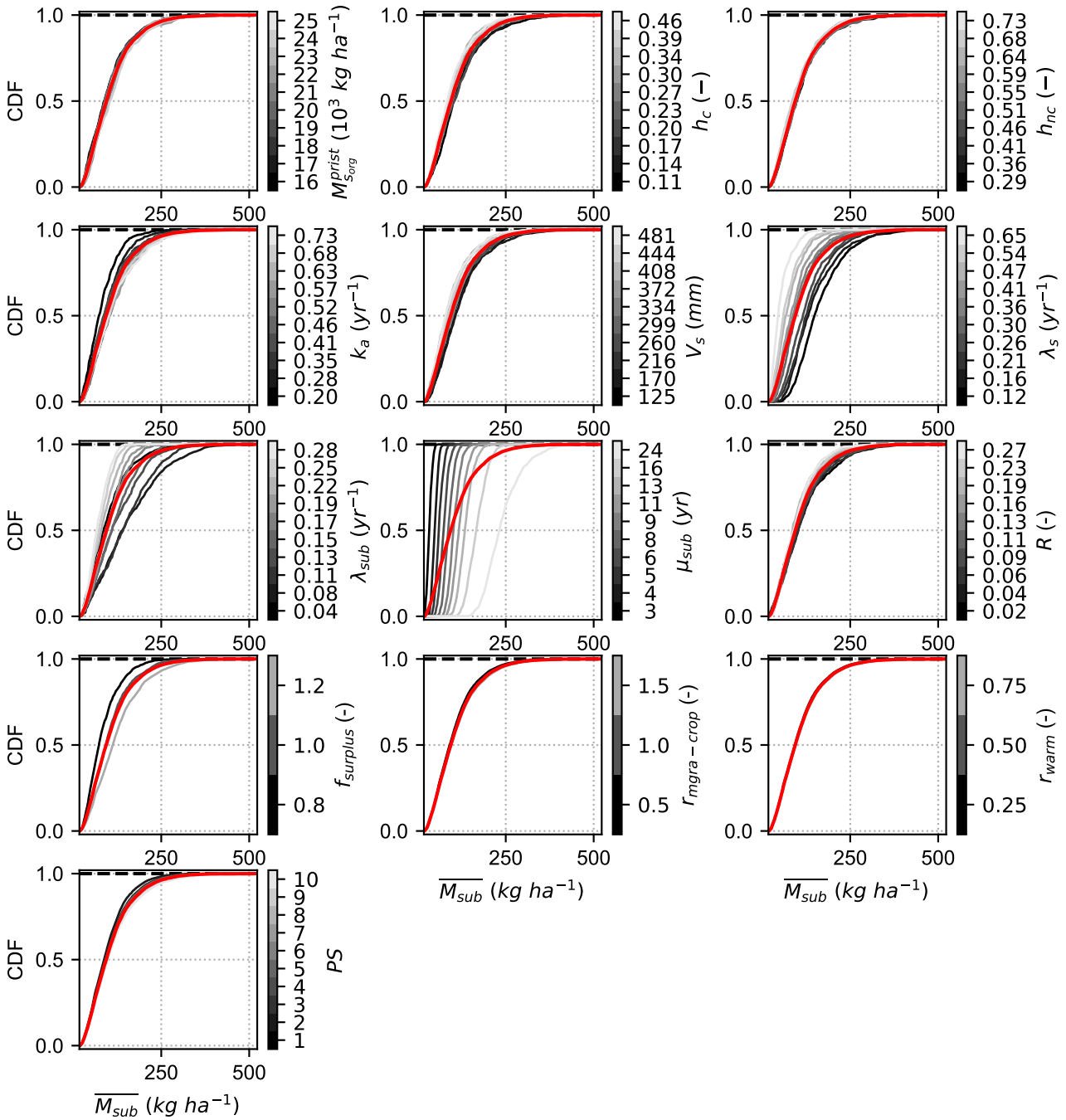

Figure S44. Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the average subsurface N storage over the period 1960–2015 ($\overline{M_{sub}}$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios and the N point sources realization (PS). These CDFs are used in the PAWN sensitivity analysis method. The colorbars report the average value of the parameters over each conditioning interval.