# 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 $\mu_{sub}$ (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 |
| $\lambda_s$ | $Q_{50}$ | 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 |
| $\lambda_{sub}$ | $Q_{50}$ | 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 |
| $\mu_{sub}$ | $Q_{50}$ | 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 |
|---|---|---|---|---|---|---|---|---|---|
| $M_{s_{org}}^{prist}$ | $\Delta Q_{50}$ | 2 | -4 | 0 | -2 | -3 | -5 | -19 | -11 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -56 | -59 | -57 | -58 | -58 | -58 | -65 | -62 |
| $h_c$ | $\Delta Q_{50}$ | -13 | -11 | -19 | 1 | -11 | 0 | -6 | -6 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -4 | -3 | -5 | 0 | -6 | -1 | 0 | -1 |
| $h_{nc}$ | $\Delta Q_{50}$ | -3 | 1 | -10 | 1 | -3 | 1 | -3 | -1 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | 1 | 1 | -0 | -0 | -1 | -1 | 1 | 0 |
| $k_a$ | $\Delta Q_{50}$ | 26 | -18 | -2 | -6 | 19 | -8 | 11 | 22 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -16 | -7 | -10 | -6 | -15 | -5 | -10 | -15 |
| $V_s$ | $\Delta Q_{50}$ | 2 | 25 | 16 | 0 | 1 | -2 | 16 | 5 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | 0 | -6 | -0 | 1 | -2 | 0 | -7 | -0 |
| $\lambda_s$ | $\Delta Q_{50}$ | -37 | -49 | -46 | -41 | -44 | -40 | -49 | -39 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -30 | -39 | -39 | -37 | -39 | -35 | -42 | -35 |
| $\lambda_{sub}$ | $\Delta Q_{50}$ | 11 | -43 | -31 | -54 | -8 | -59 | 1 | 2 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -13 | -12 | -19 | -21 | -8 | -16 | -12 | -8 |
| $\mu_{sub}$ | $\Delta Q_{50}$ | -63 | -33 | -47 | -37 | -70 | -27 | -59 | -68 |
| (%) | $\Delta(Q_{97.5} - Q_{2.5})$ | -41 | -31 | -34 | -20 | -56 | -11 | -39 | -53 |
| $R$ | $\Delta Q_{50}$ | -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 ($\Delta Q_{50}$) 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

| | $M_{s_{org}}^{prist}$ | $h_c$ | $h_{nc}$ | $k_a$ | $V_s$ | $\lambda_s$ | $\lambda_{sub}$ | $\mu_{sub}$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|
| $M_{s_{org}}^{prist}$ | 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 |
| $\lambda_s$ | 0.01 | **-0.06** | -0.01 | -0.02 | **-0.09** | 1.0 | **-0.36** | **-0.54** | **-0.08** |
| $\lambda_{sub}$ | 0.01 | -0.02 | -0.02 | **-0.13** | 0.03 | **-0.36** | 1.0 | **-0.32** | -0.02 |
| $\mu_{sub}$ | 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 $\lambda_s$ and the denitrification rate constant in the subsurface $\lambda_{sub}$ (-0.36), between $\lambda_s$ and the mean travel time in the subsurface $\mu_{sub}$ (-0.54), between $\mu_{sub}$ and $\lambda_{sub}$ (-0.32), between $\mu_{sub}$ 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 |
| $\Delta M_s$ | $Q_{50}$ | 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 |
| $\Delta M_s$ | $Q_{50}$ | 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 |
| $\Delta M_p$ | $Q_{50}$ | 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 |
| $\Delta M_a$ | $Q_{50}$ | 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 |
| $\Delta M_{in}$ | $Q_{50}$ | -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; $\Delta M_s$, $\Delta M_p$, $\Delta M_a$, and $\Delta M_{in}$ 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.

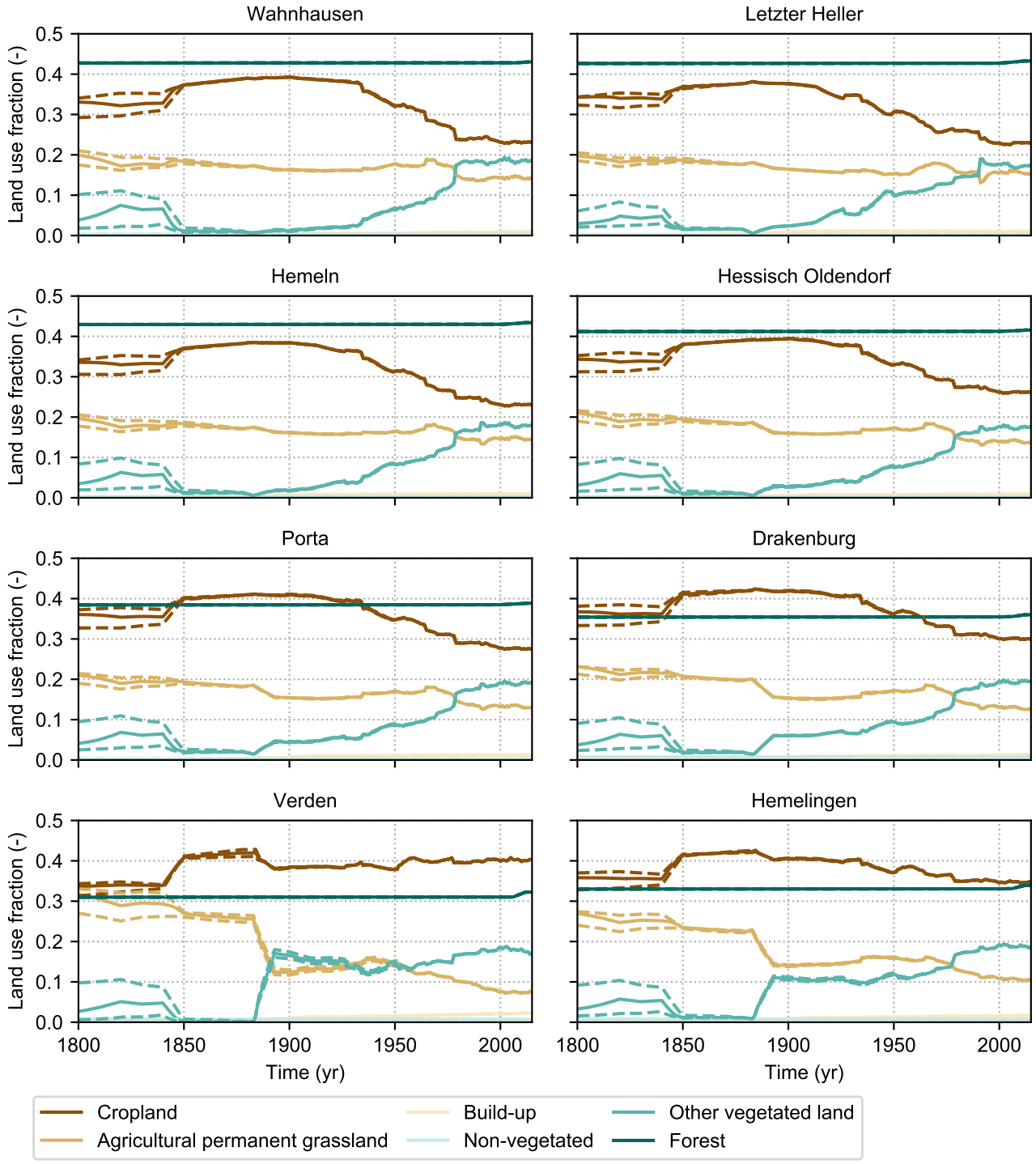


**Figure S1.** Time series of the land use fractions, i.e., cropland, agricultural permanent grassland, built-up areas, non-vegetated areas, other vegetated land (which includes in particular natural grassland, urban parks, and green areas in discontinuous urban fabric) and forest for the period 1800–2015 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.

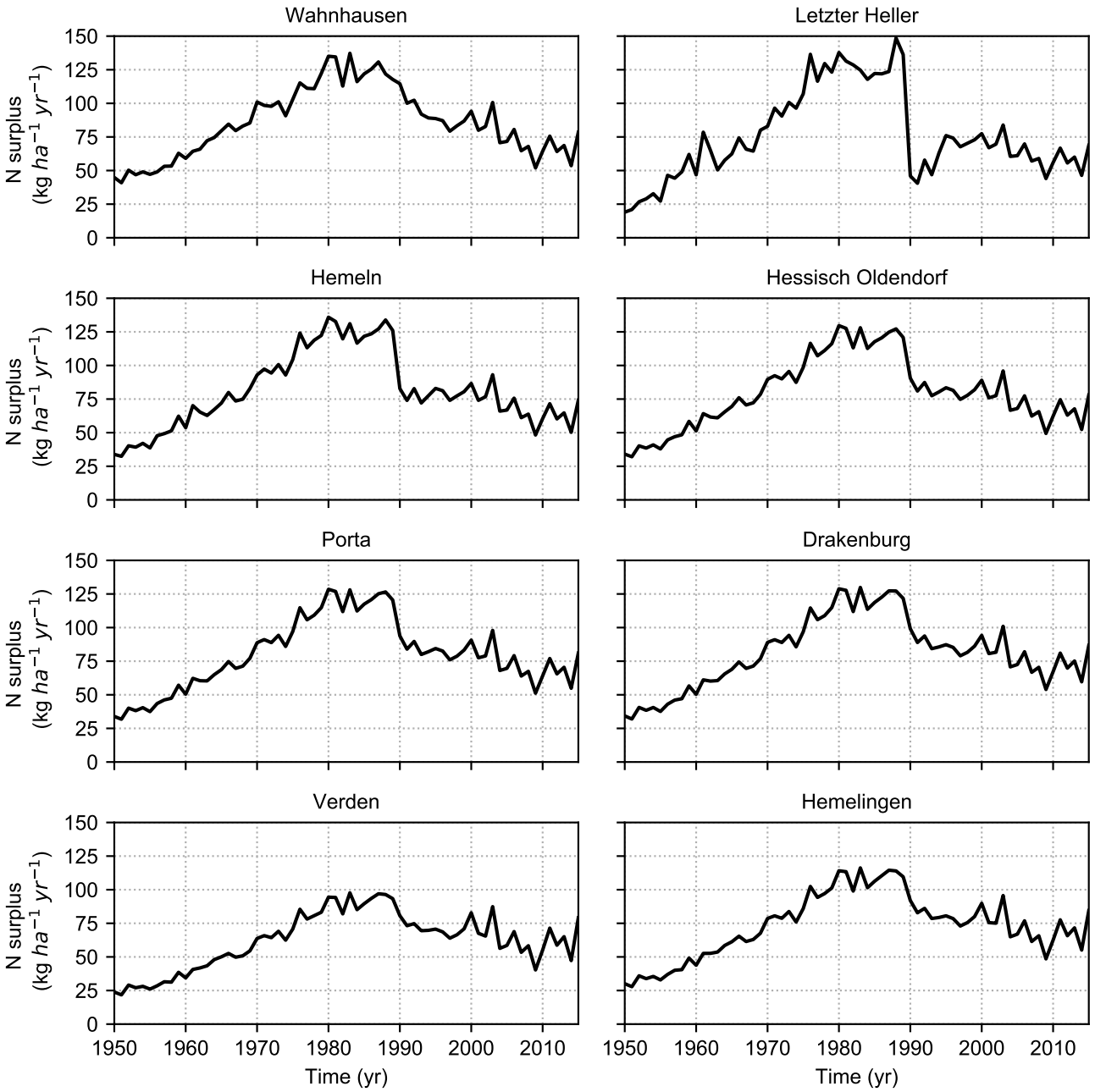


**Figure S2.** Time series of the agricultural N surplus **(per unit of agricultural area)** for the period 1950–2015 for the eight subcatchments.

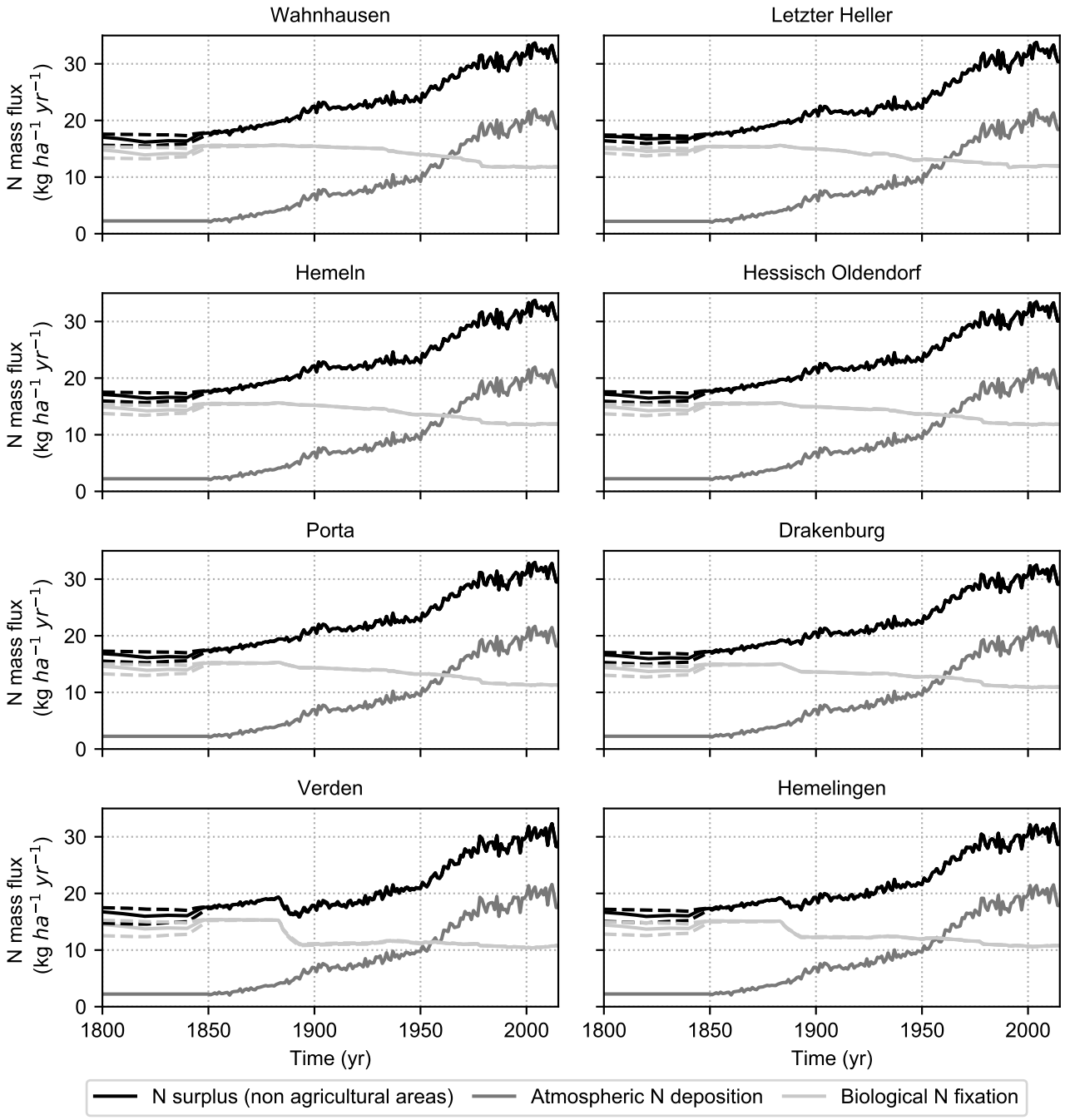


**Figure S3.** Time series of the total N surplus for non-agricultural areas and its components i.e., atmospheric N deposition and biological N fixation **(per unit of non-agricultural areas)** for the period 1800–2015 for the eight subcatchments. For biological N fixation, the solid lines are obtained 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.

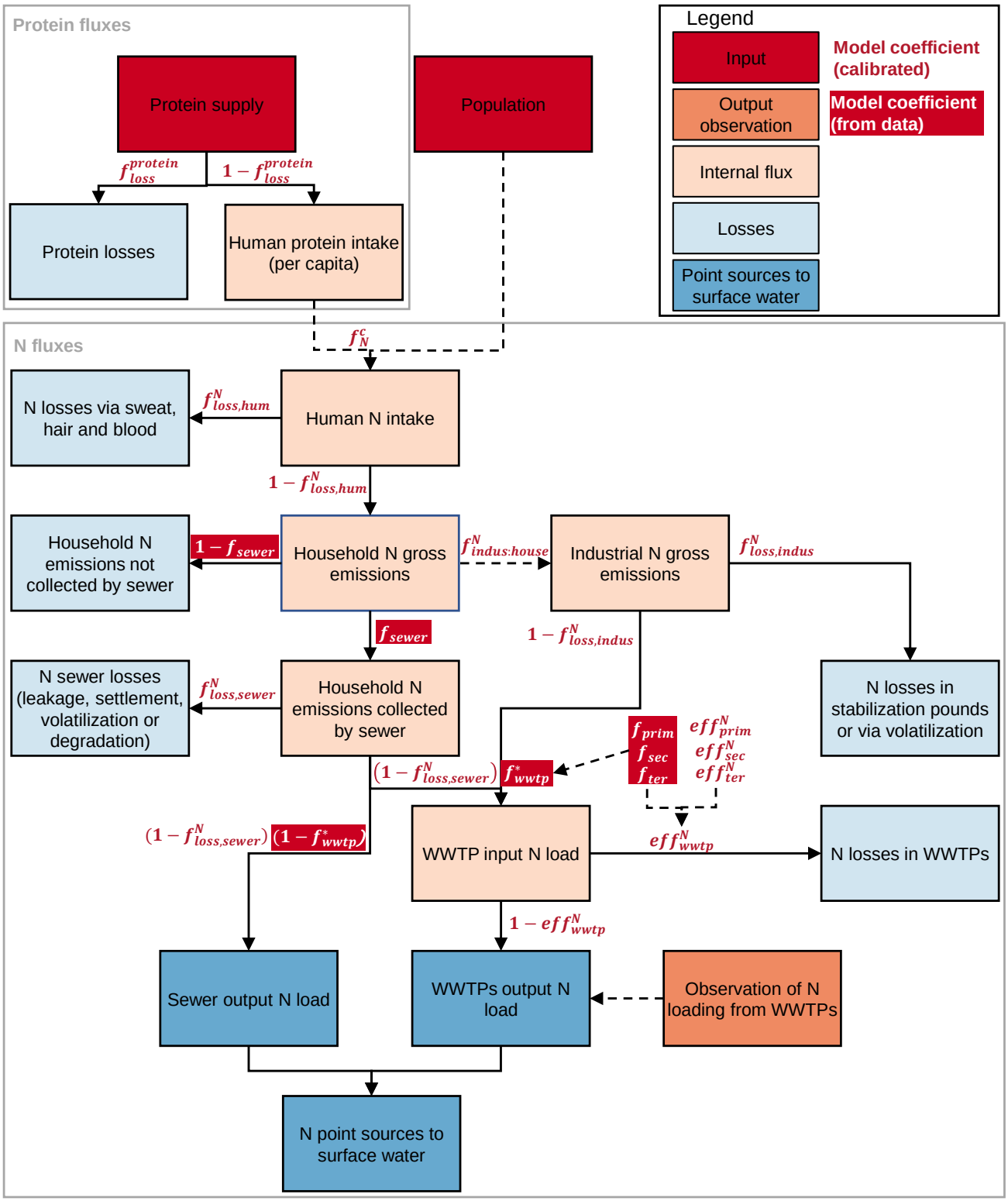


**Figure S4.** Flow chart depicting the procedure for the calculation of the N point sources. Details on the datasets used are reported in Table S1, and calibrated model coefficients are described in Table S2. WWTP: Waste Water Treatment Plant.

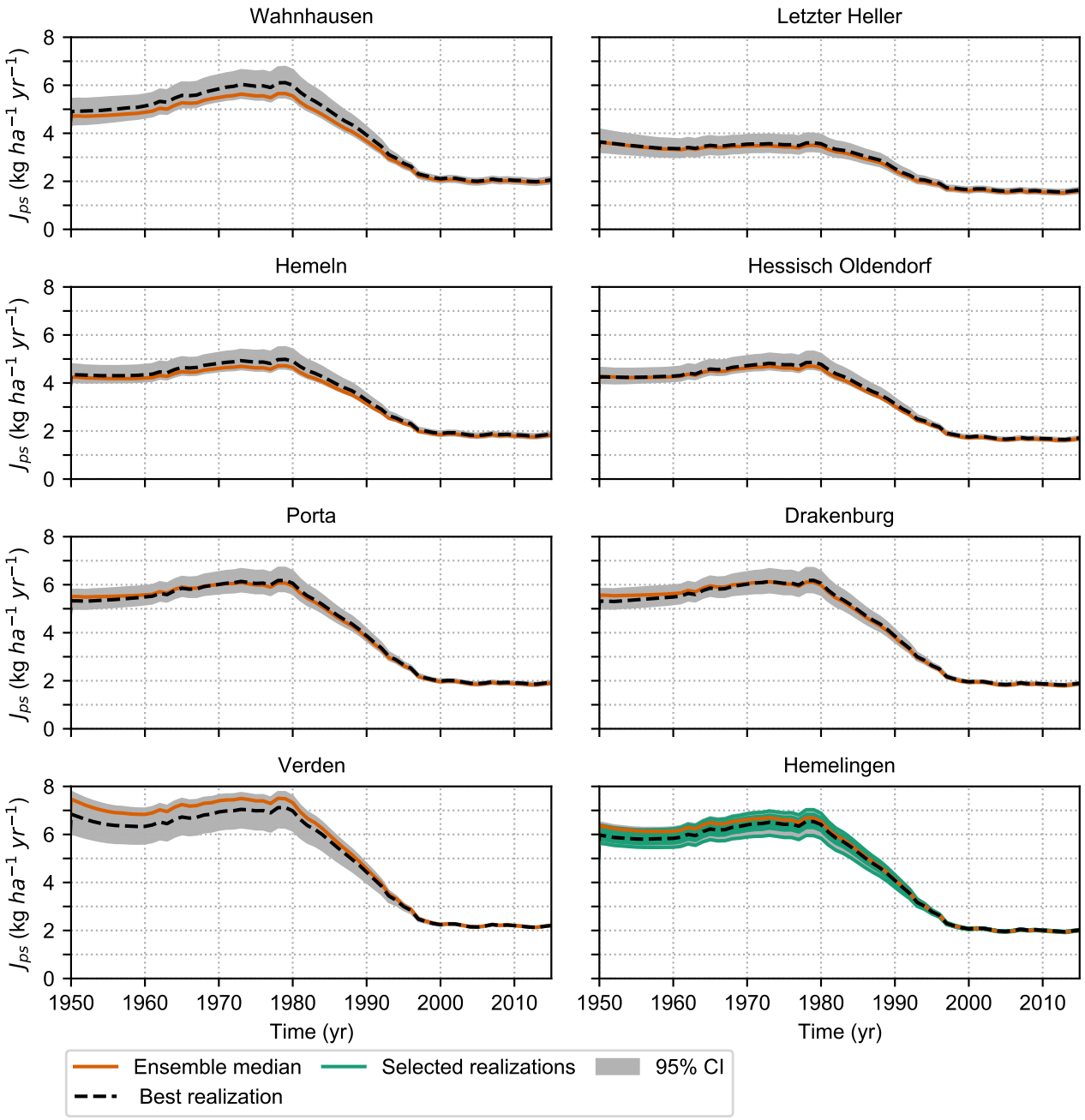


**Figure S5.** Time series of the N point sources for the period 1950–2015 for the eight subcatchments. The dashed black lines represent the baseline realization that we select to estimate the model parameters for the eight subcatchments (Section 4.1). This is the realization that presents the smallest 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).

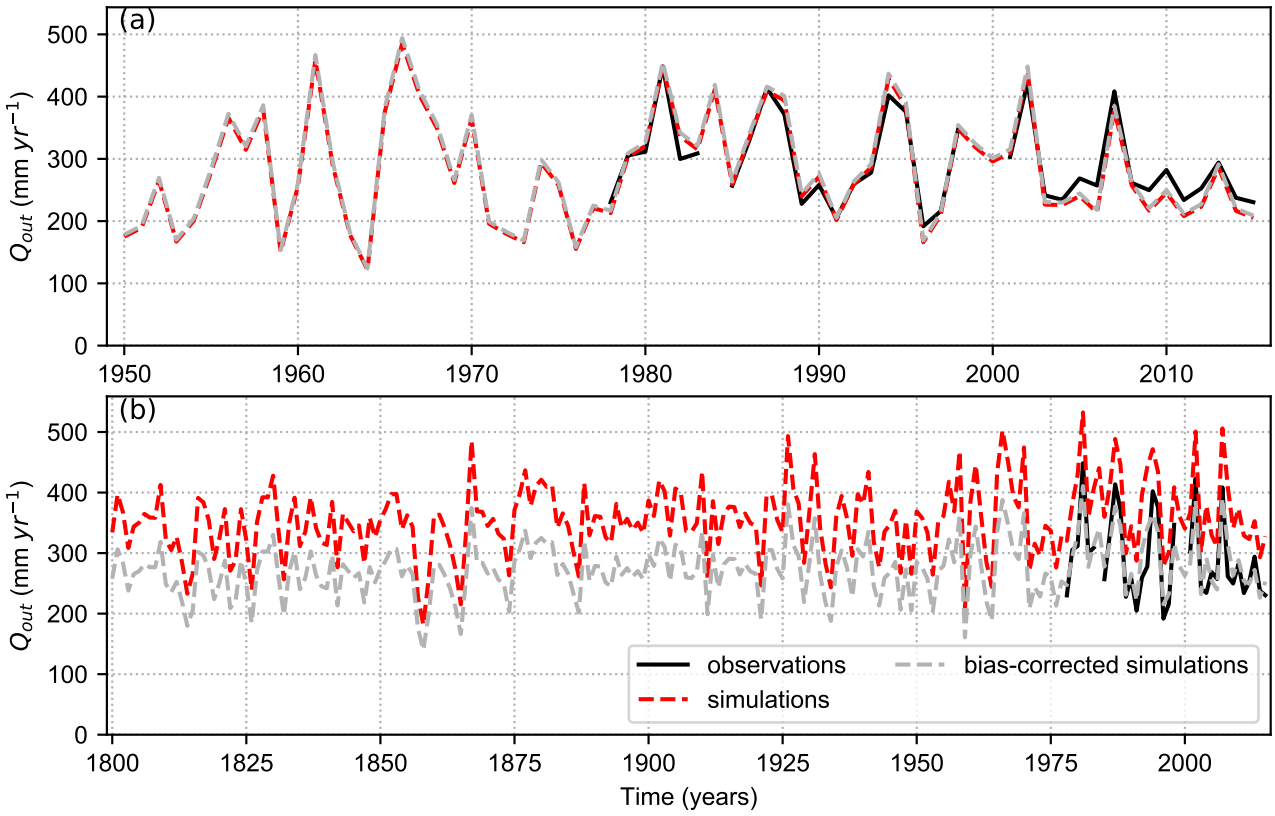


**Figure S6.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Bonaforth** station (Wahnhausen subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

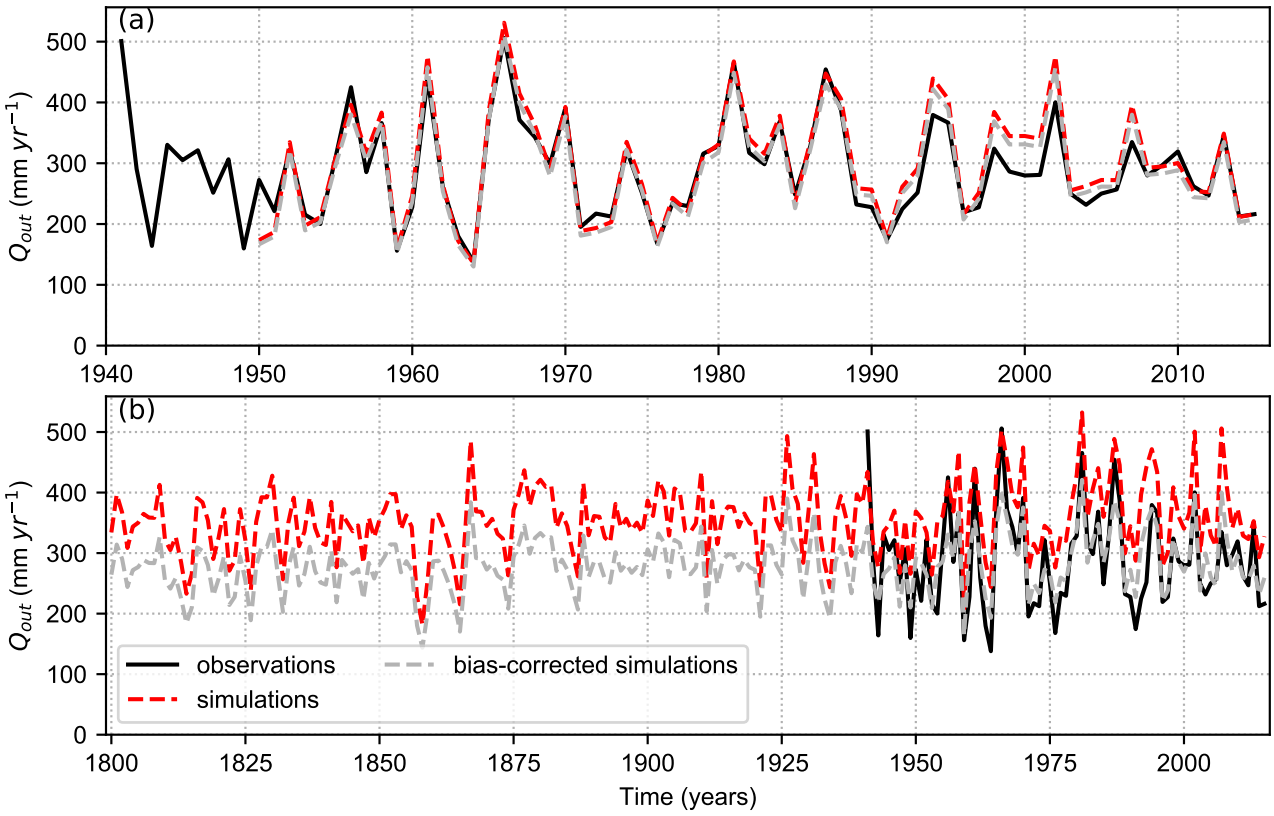


**Figure S7.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Letzter Heller** station (Letzter Heller subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

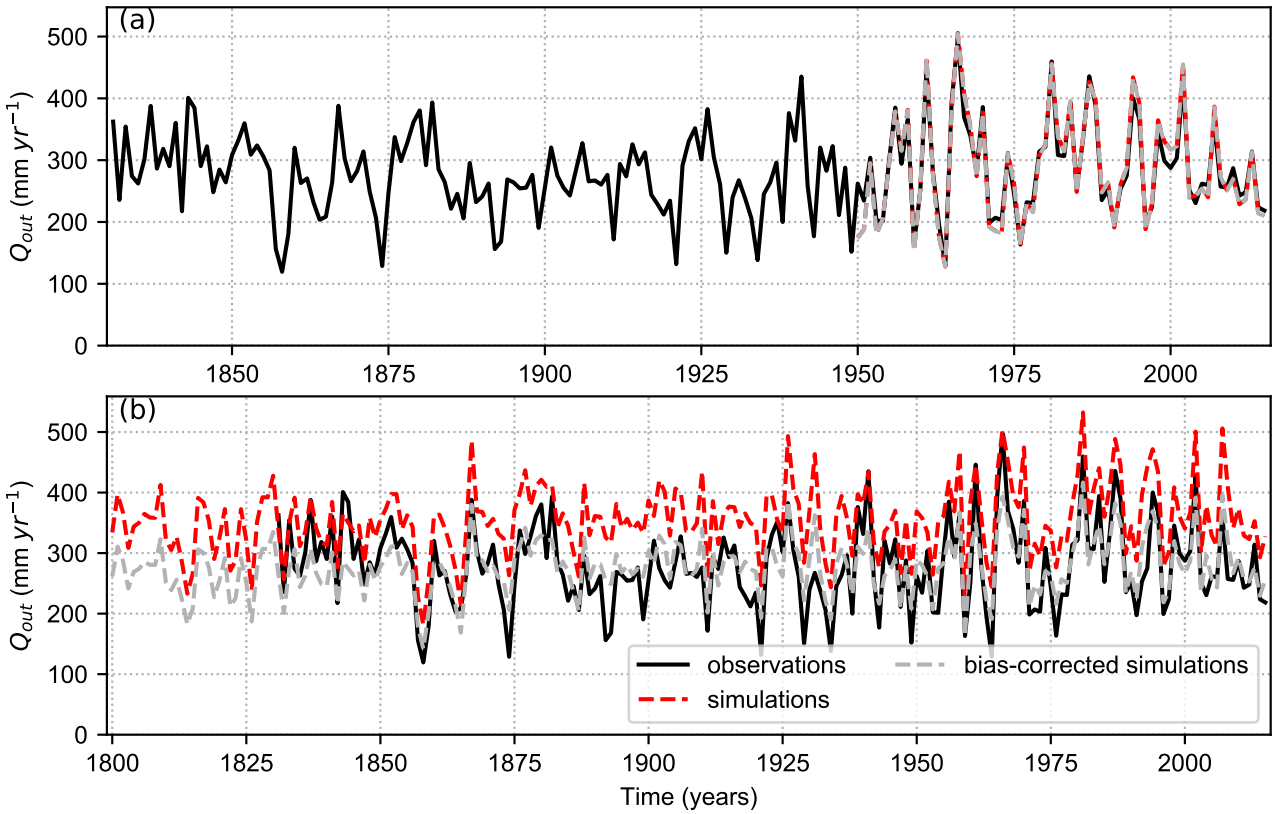


**Figure S8.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Hann. Münden** station (Hemeln subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

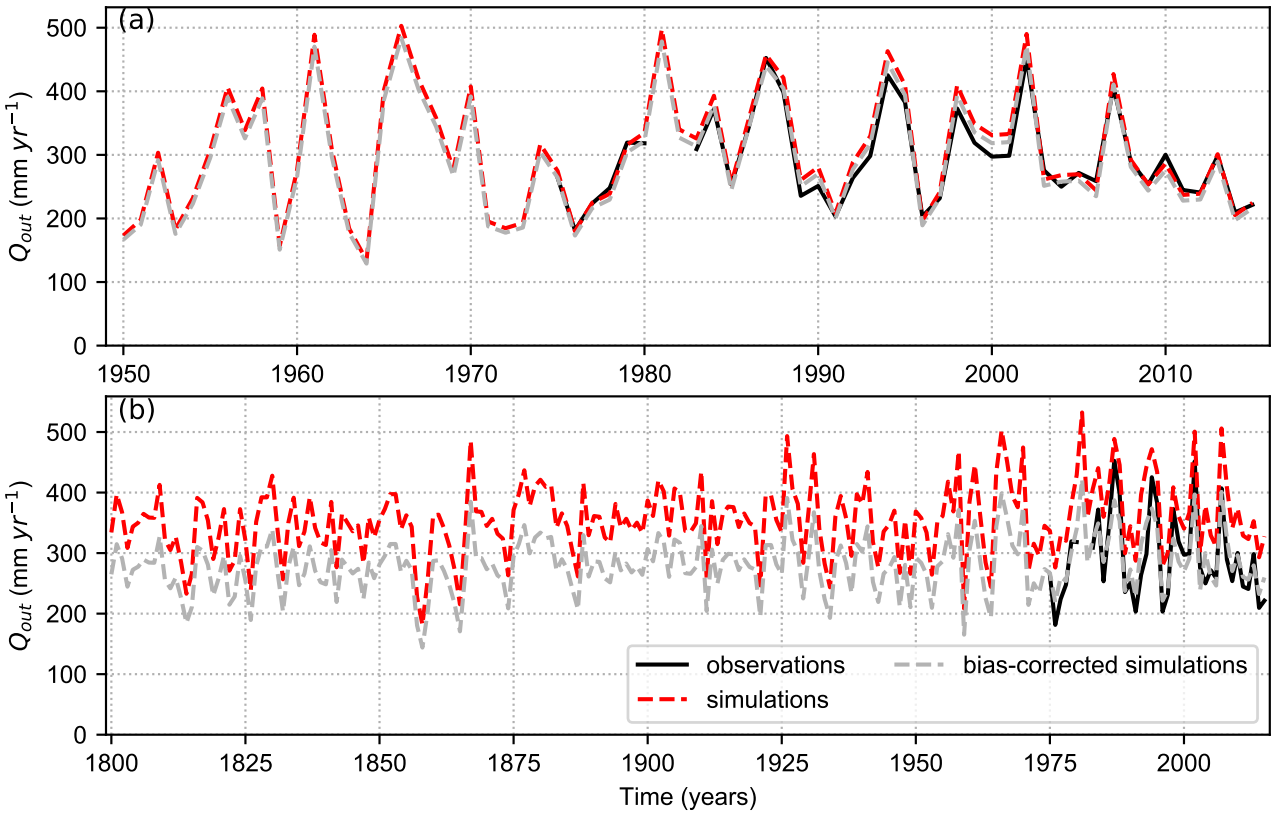


**Figure S9.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Hameln** station (Hessisch Oldendor subcatchment): (a) Medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

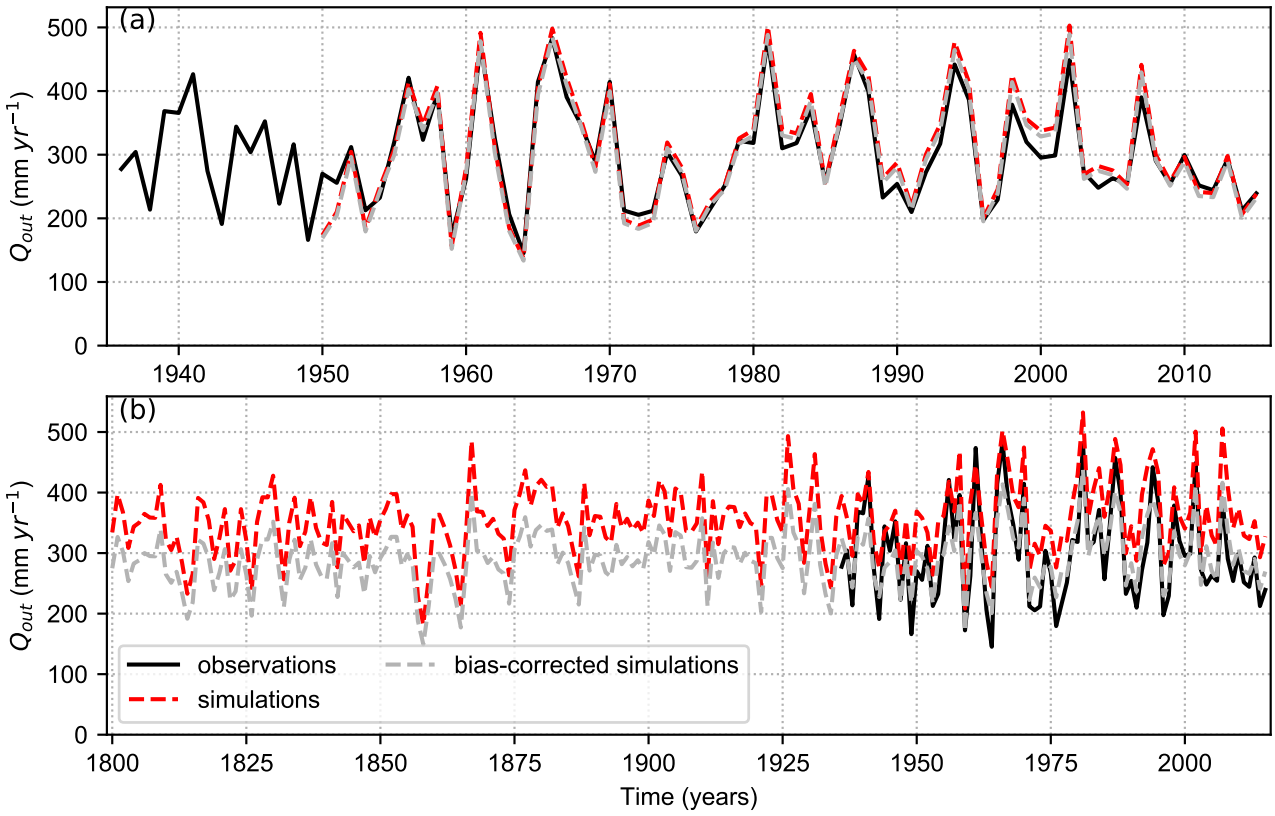


**Figure S10.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Porta** station (Porta subcatchment): (a) medium-term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

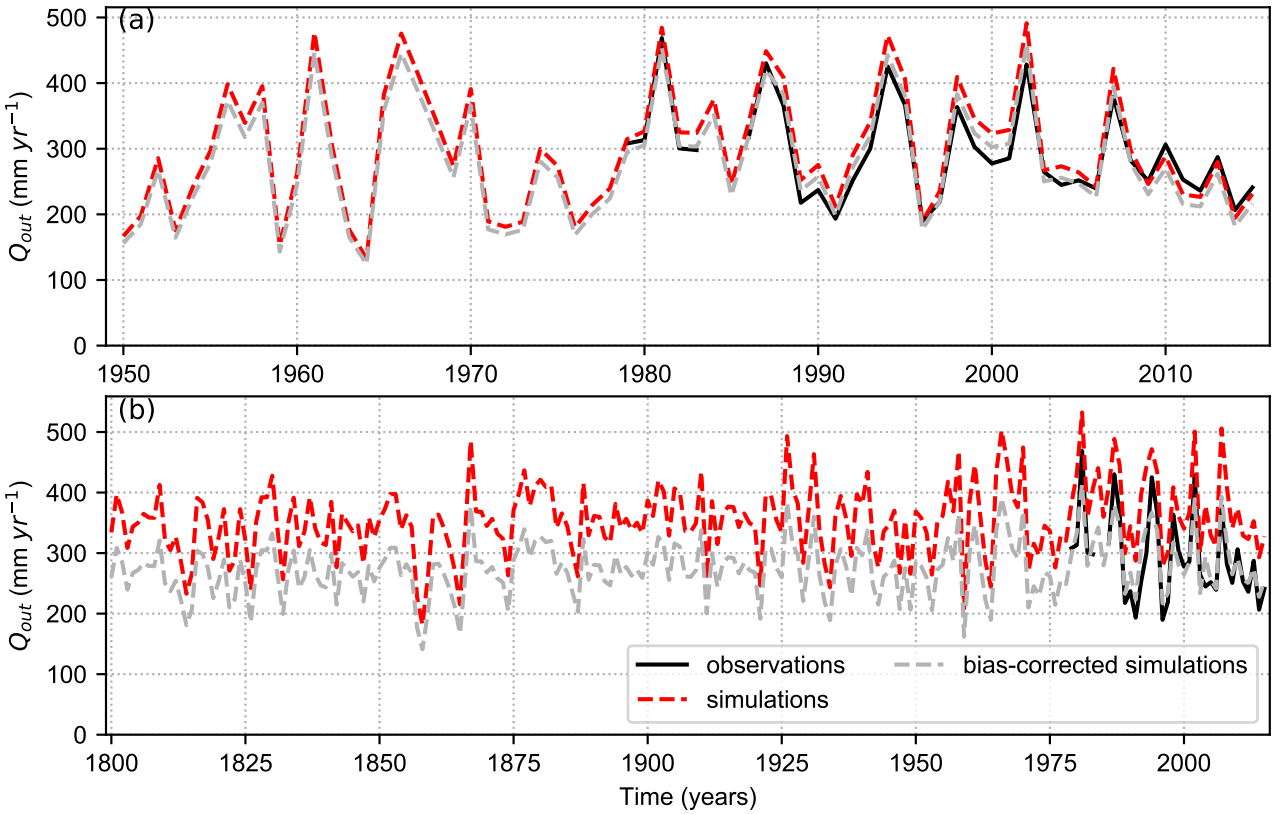


**Figure S11.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Drakenburg** station (Drakenburg subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

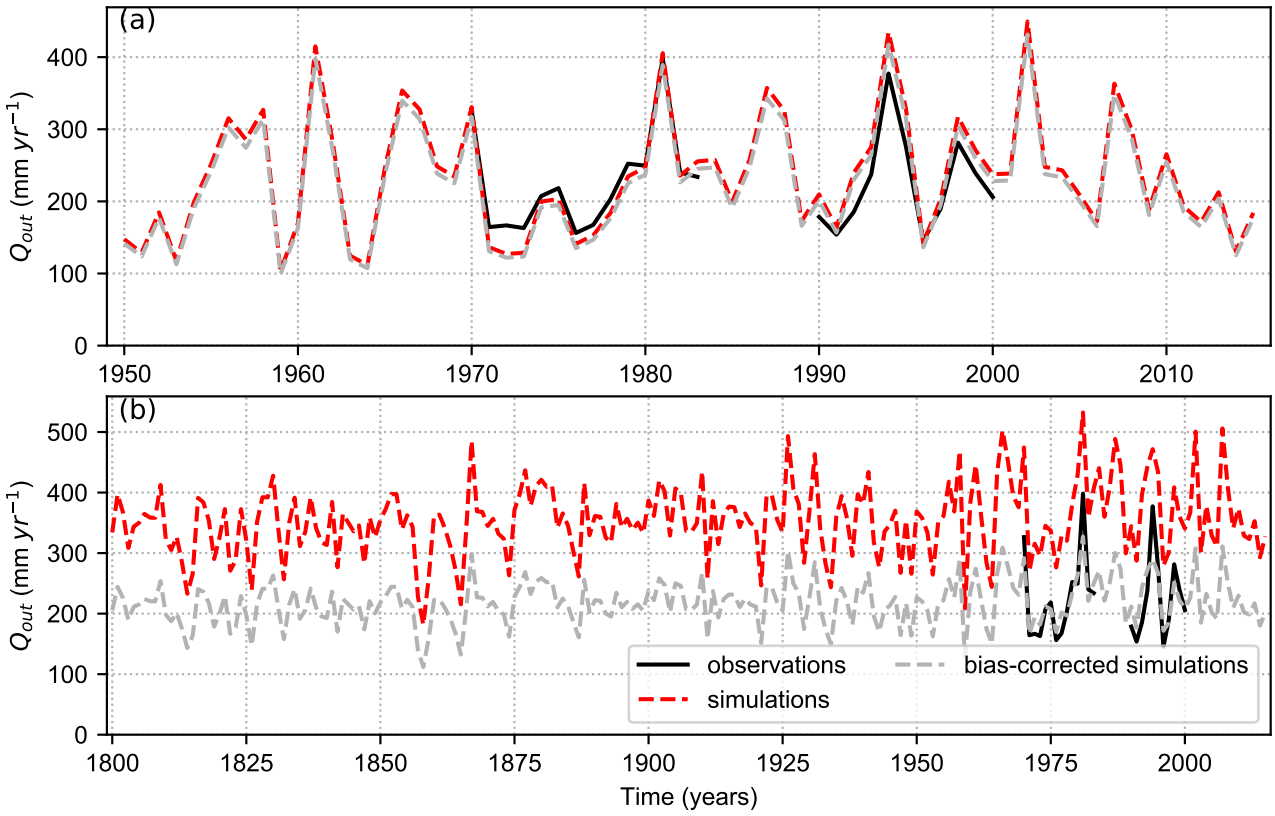


**Figure S12.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Westen** station (Verden subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

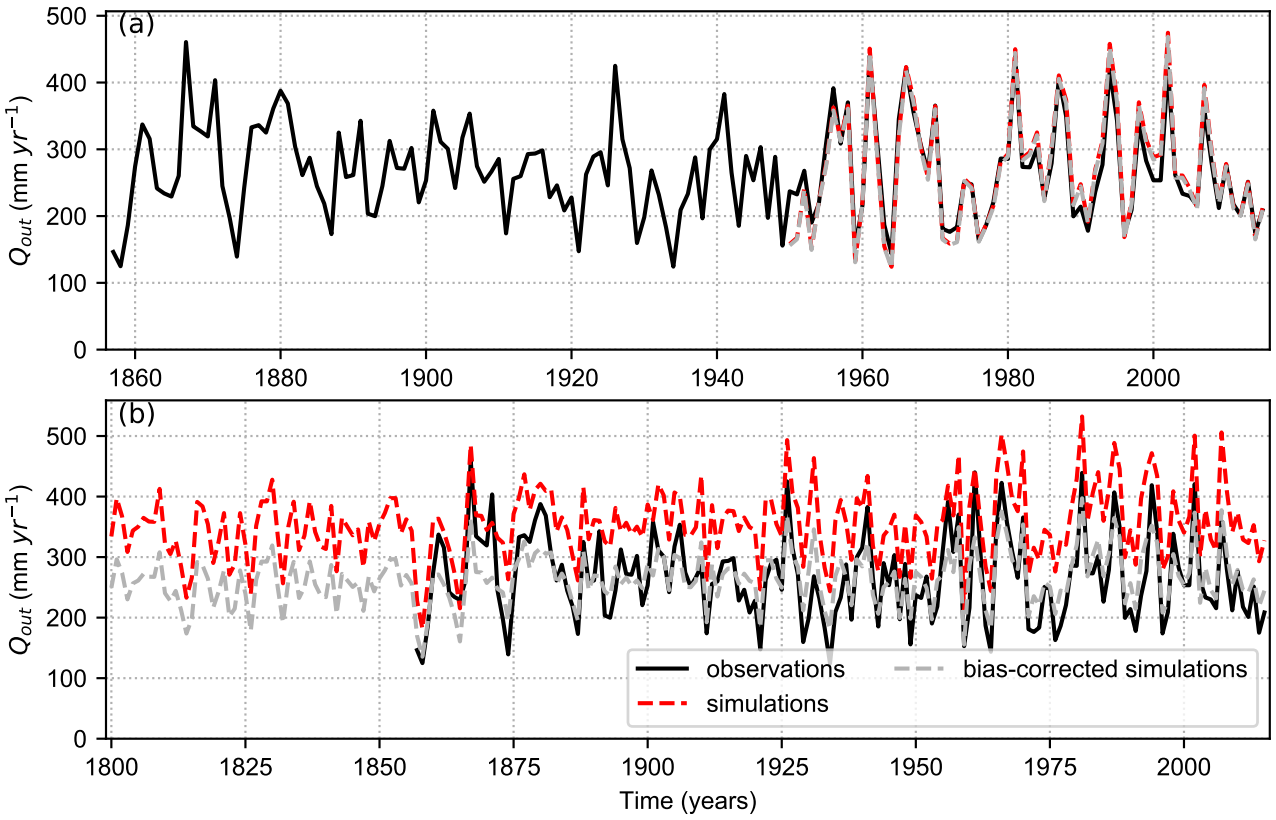


**Figure S13.** Observed, simulated and bias-corrected simulated annual discharge $Q_{out}$ for the **Intschede** station (Hemelingen subcatchment): (a) Medium term simulations (Zink et al., 2017) and (b) Long-term simulations (Hanel et al., 2018).

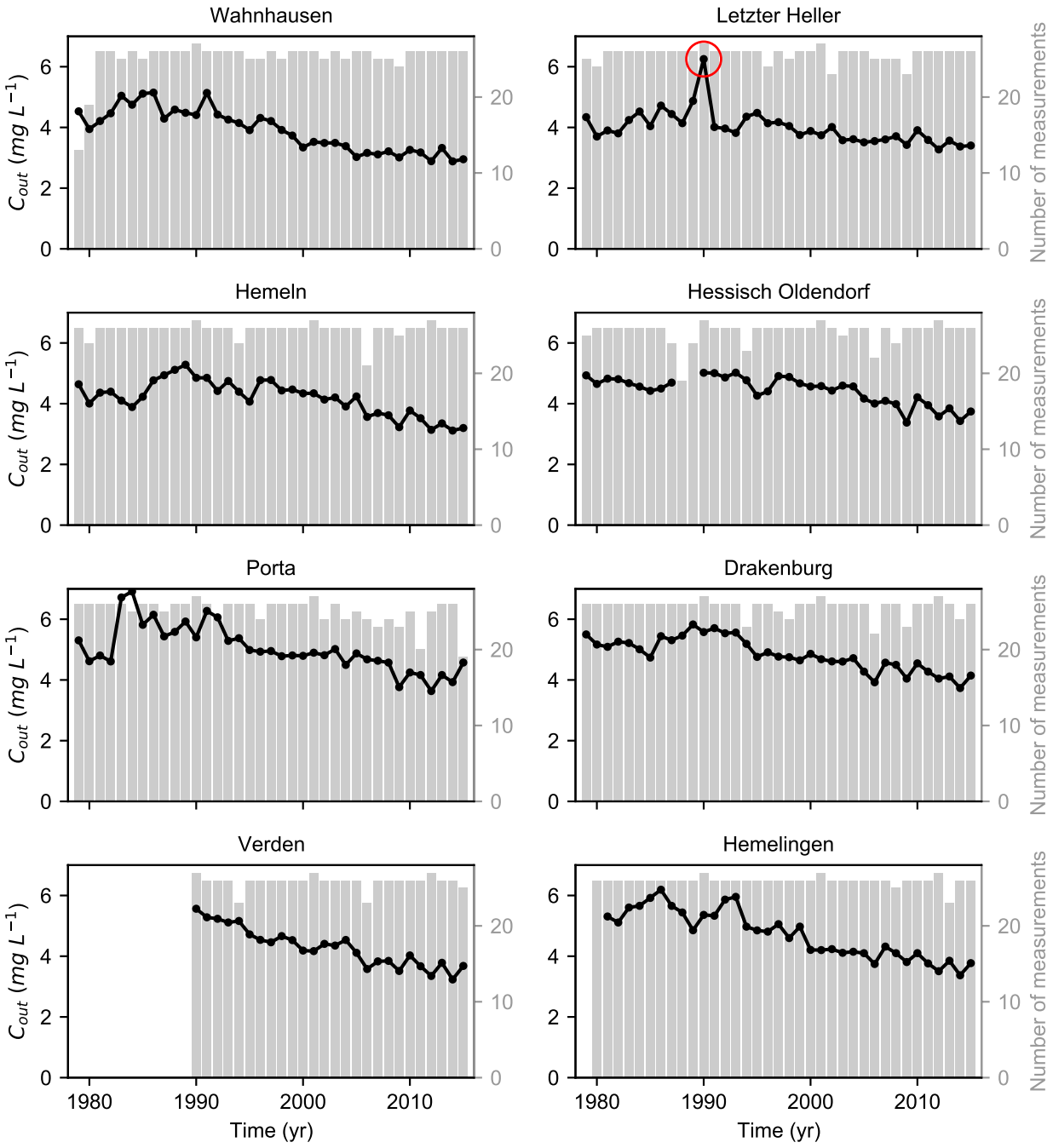


**Figure S14.** Time series of annual observed in-stream N-$NO_3$ concentration at the catchment outlet $C_{out}$ and number of 14-day average measurements provided for each year by the River Basin Commission Weser (FGG Weser, 2021). For the Letzter Heller catchments, we combine the concentration measurements 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.

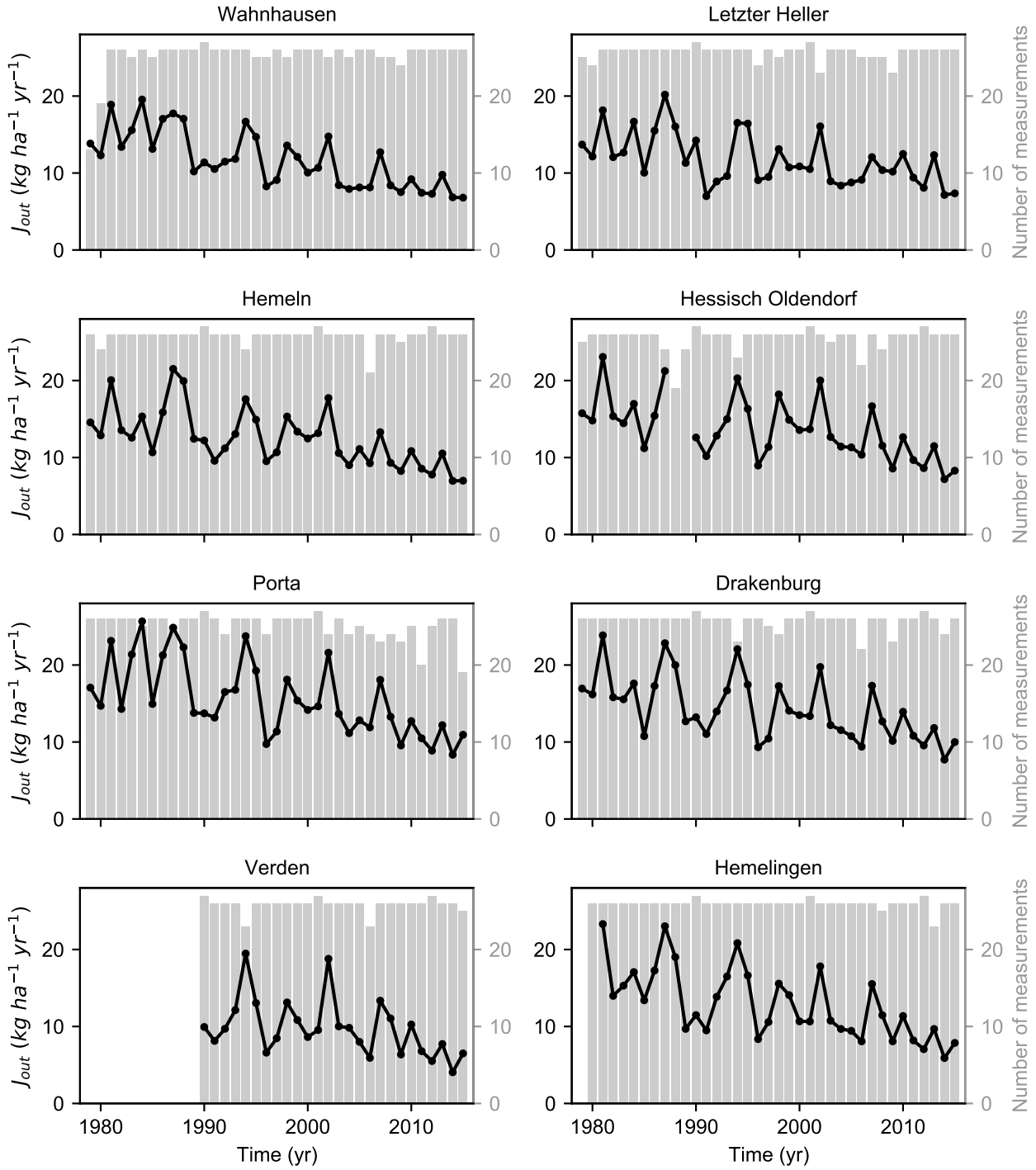


**Figure S15.** Time series of annual average observed in-stream N-$NO_3$ loading at the catchment outlet $J_{out}$ and number of 14-day average measurements provided for each year by the River Basin Commission Weser (FGG Weser, 2021). For the Letzter Heller catchments, we combine the concentration measurements 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.

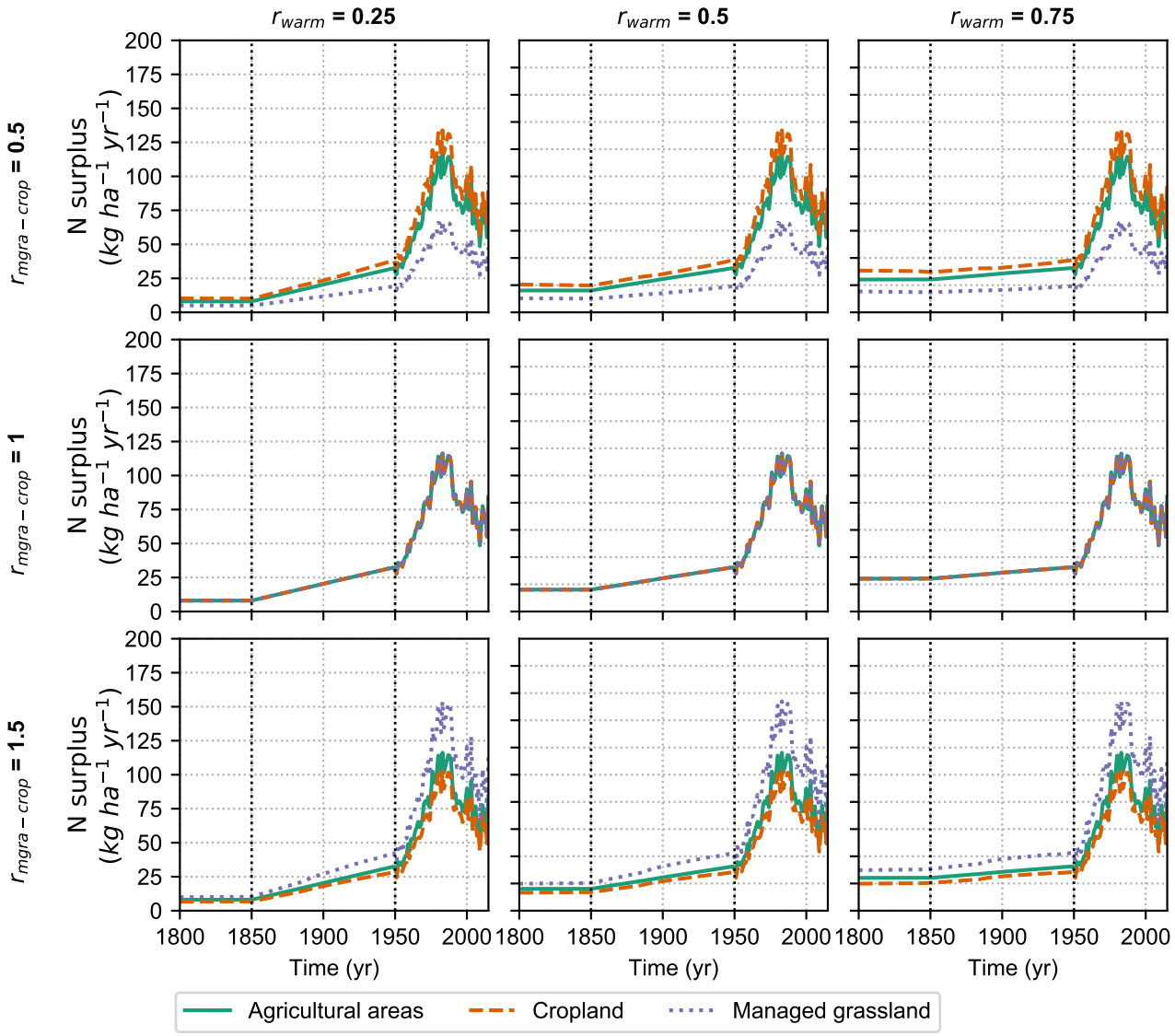


**Figure S16.** N surplus for total agricultural areas, cropland and agricultural permanent grassland for the period 1800–2015 for nine scenarios defined in this study for **Hemelingen**, corresponding to a value of the N surplus multiplier ($f_{surplus}$) equal to 1. The figure reports the corresponding 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 $r_{warm} = 0.5$).

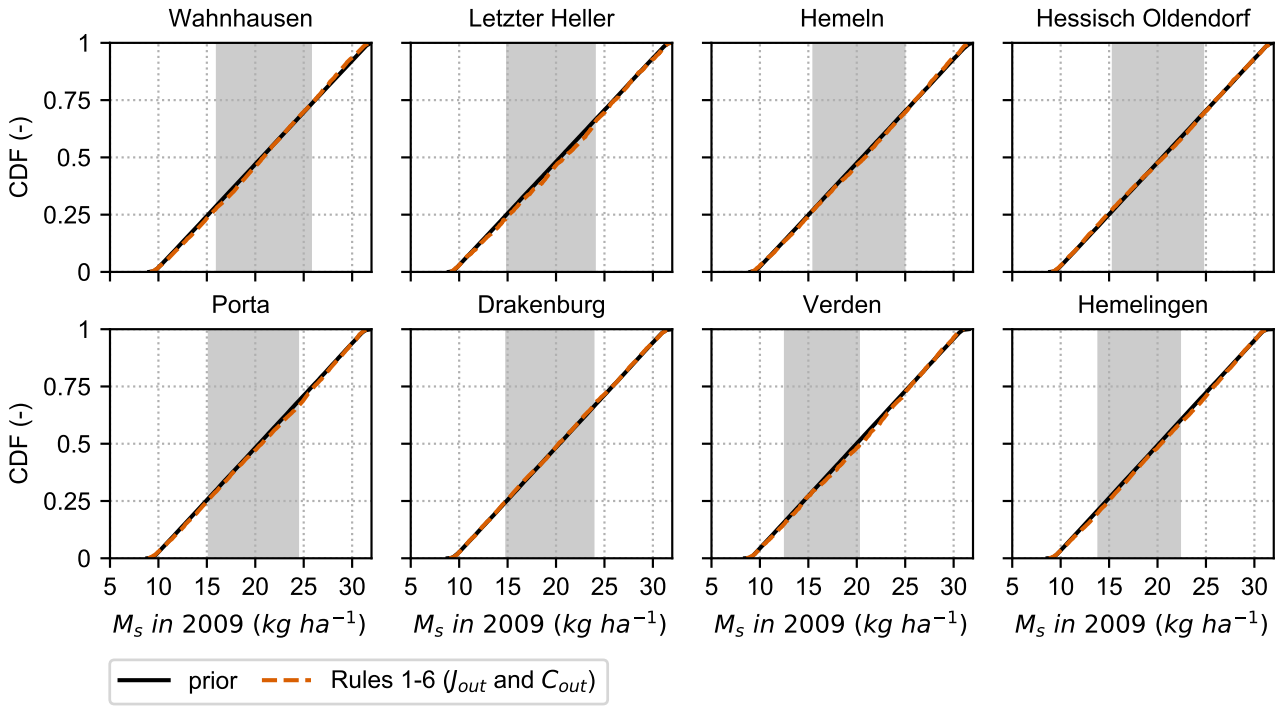


**Figure S17.** Cumulative Distribution Function (CDF) of the source zone N content $M_s$ in 2009 in the initial simulation ensemble (100,000 realizations) and after application of rules 1-6 (rules on the in-stream loading $J_{out}$ and concentration $C_{out}$) for the eight subcatchments. The grey shaded 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).

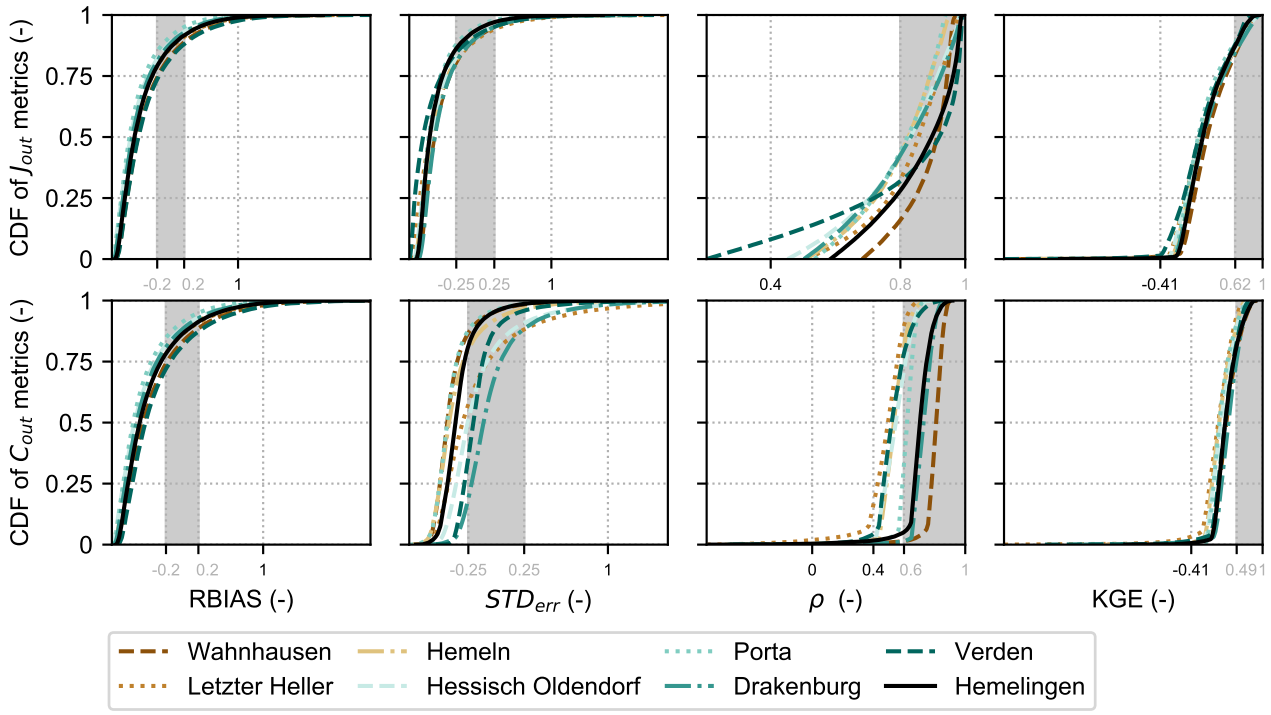


**Figure S18.** Application of the soft rules for the eight subcatchments: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream N loading ($J_{out}$) and concentration ($C_{out}$) in the initial simulation ensemble (100,000 realizations). The figure reports the three performance metrics used in the definition of rules 1-6 (the relative bias *RBIAS*, the variability error $STD_{err}$ and the Pearson correlation coefficient $\rho$) 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.

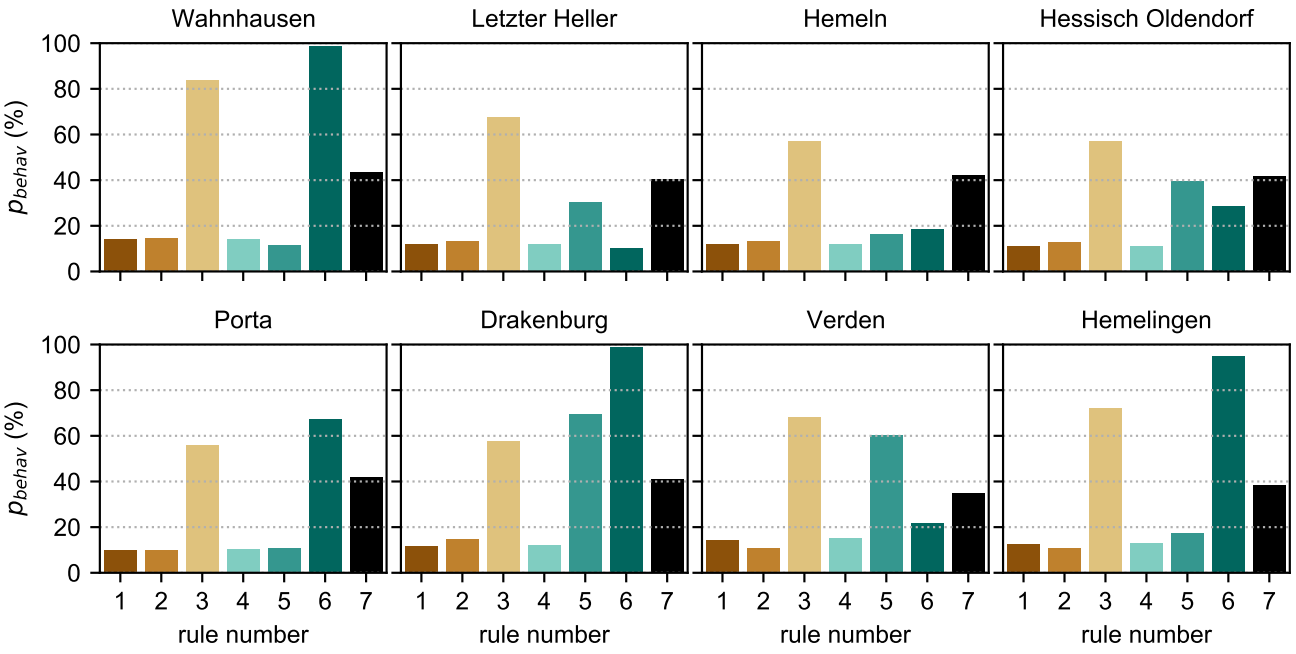


**Figure S19.** Percentage of behavioural model realizations when applying each of the seven rules individually for the eight subcatchments.

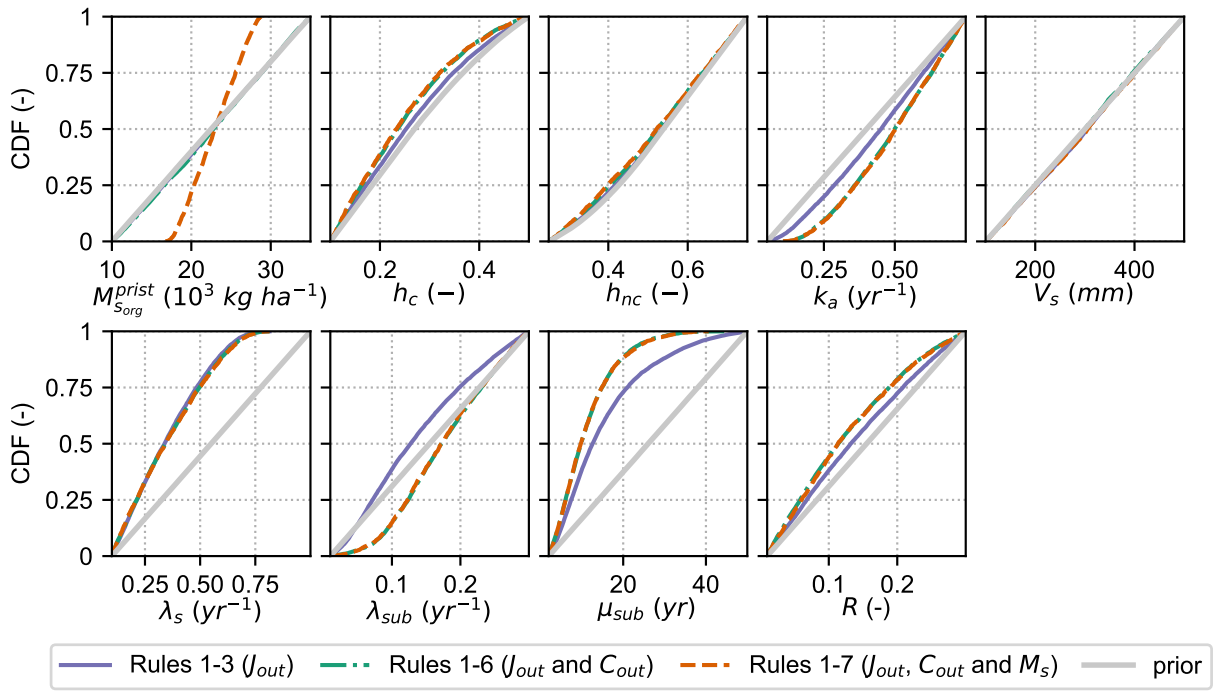


**Figure S20.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Wahnhausen**.

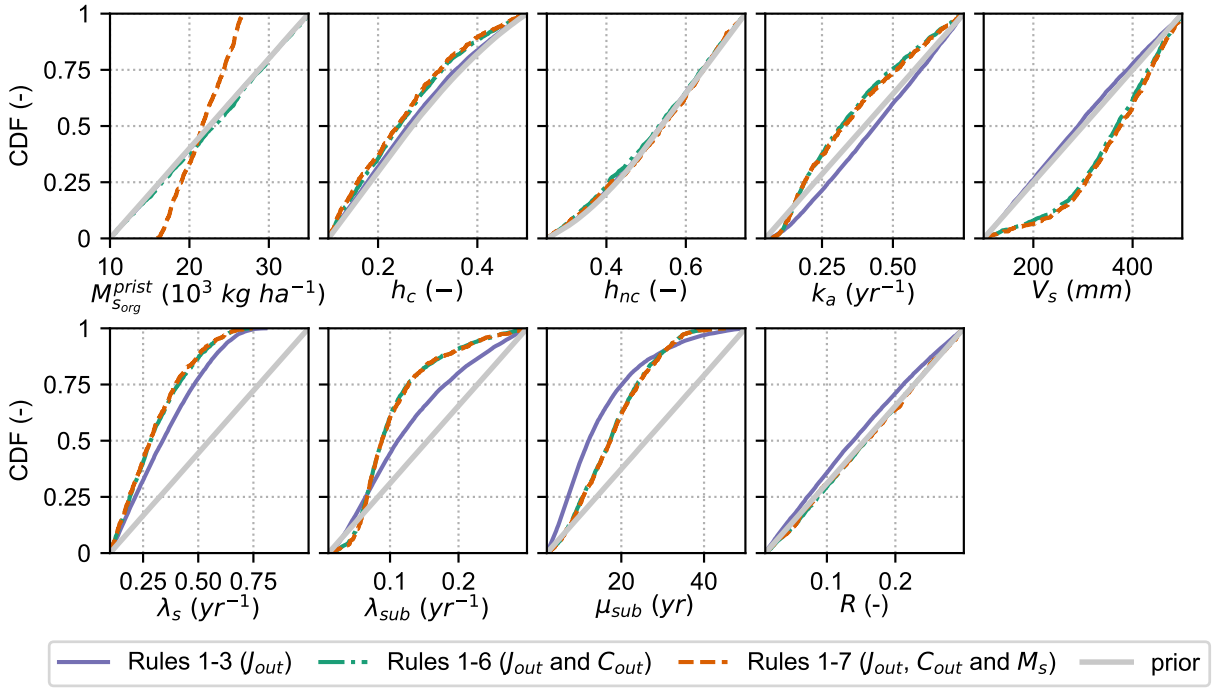


**Figure S21.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Letzter Heller**.

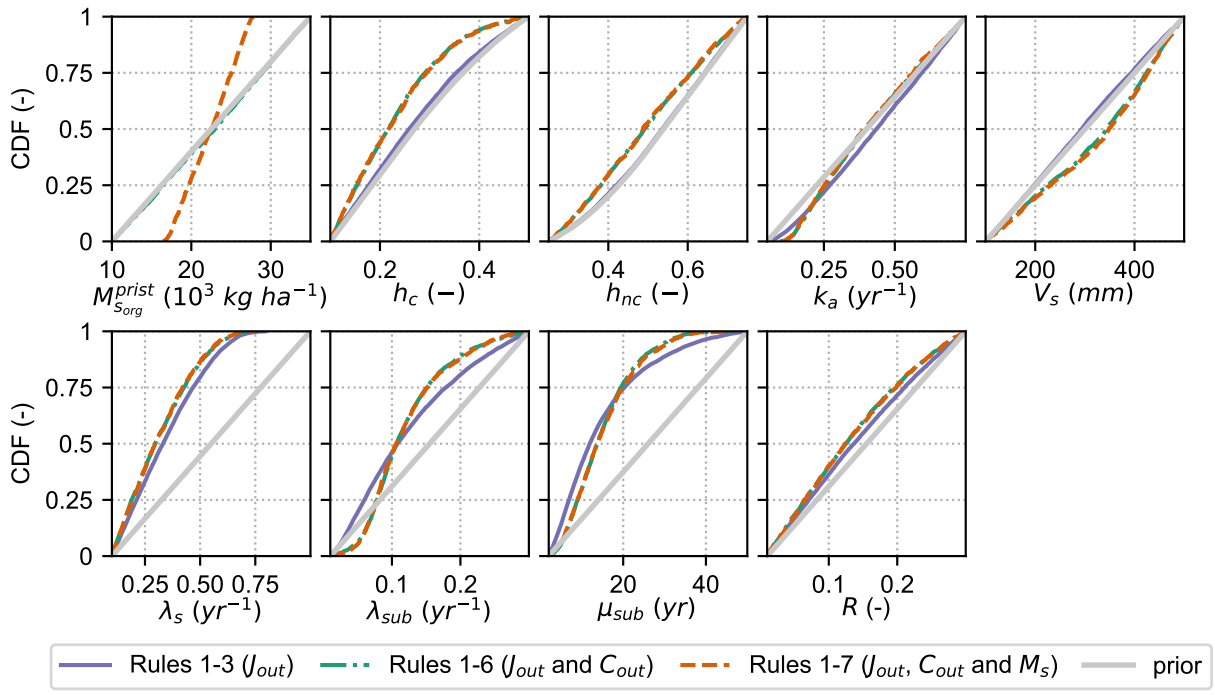


**Figure S22.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hemeln**.

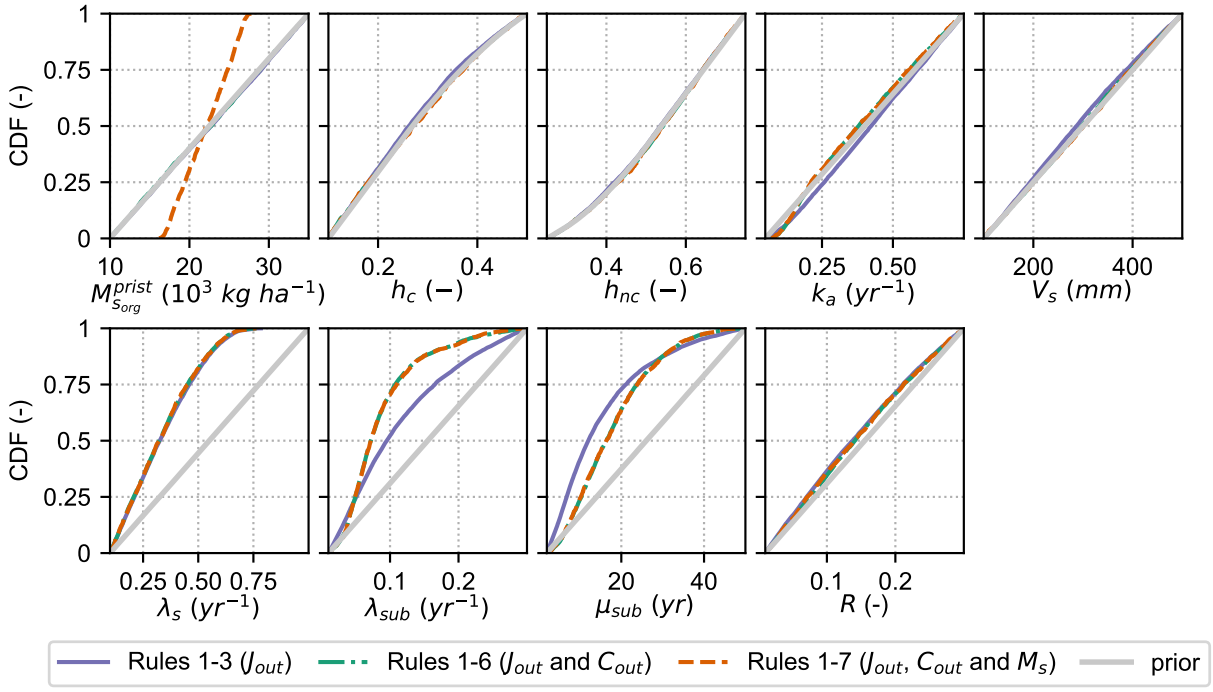


**Figure S23.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hessisch Oldendorf**.

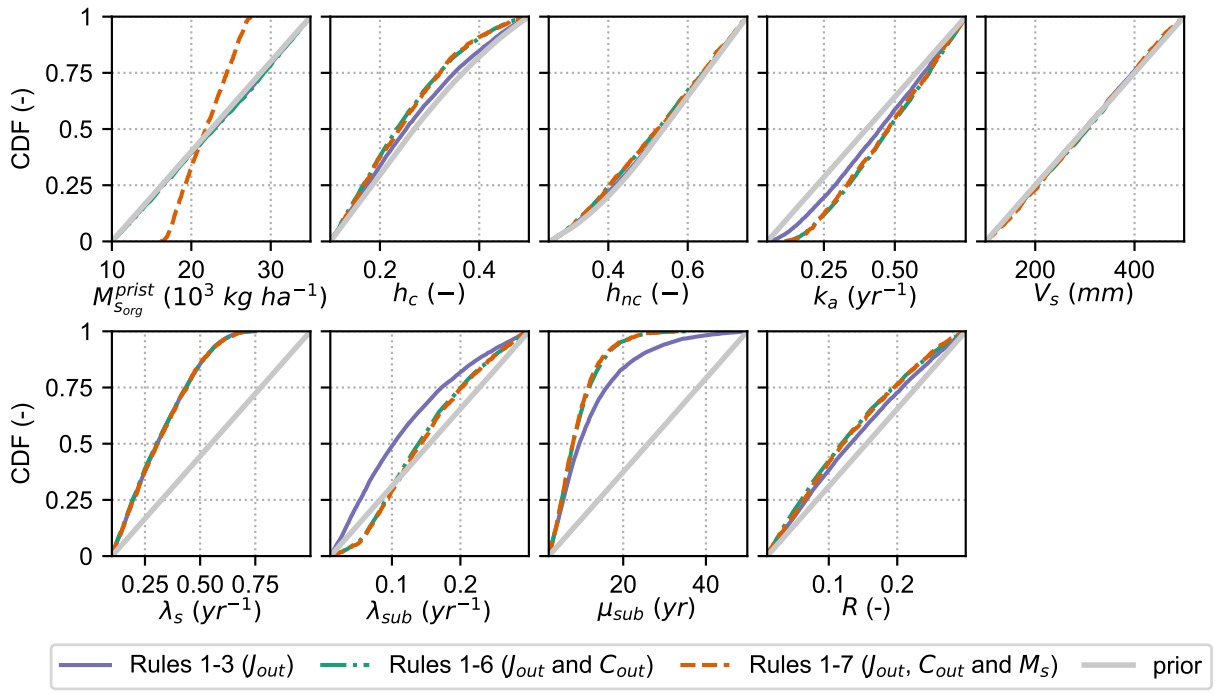


**Figure S24.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Porta**.

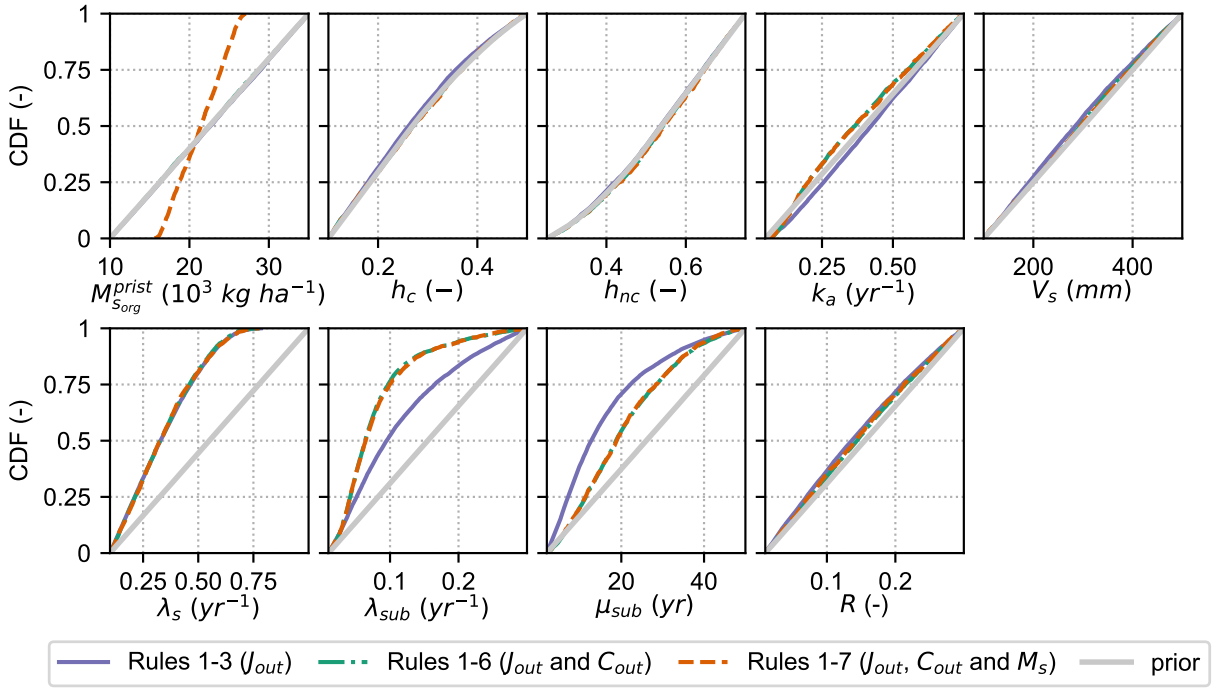


**Figure S25.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Drakenburg**.

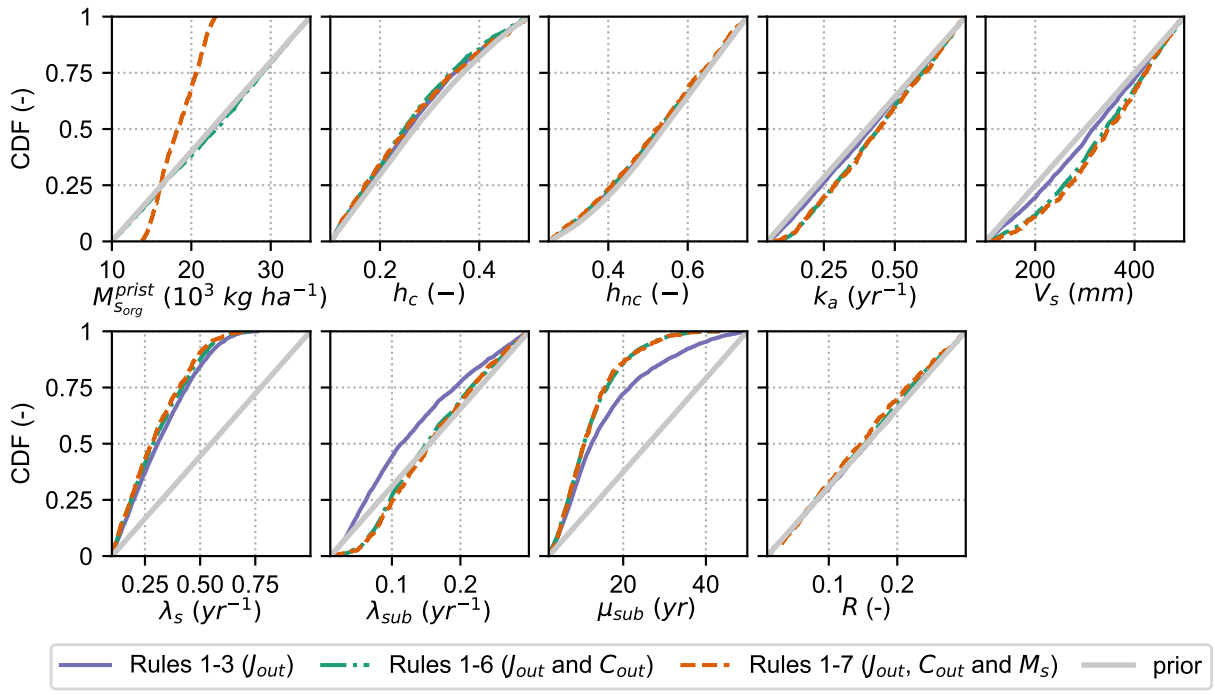


**Figure S26.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Verden**.

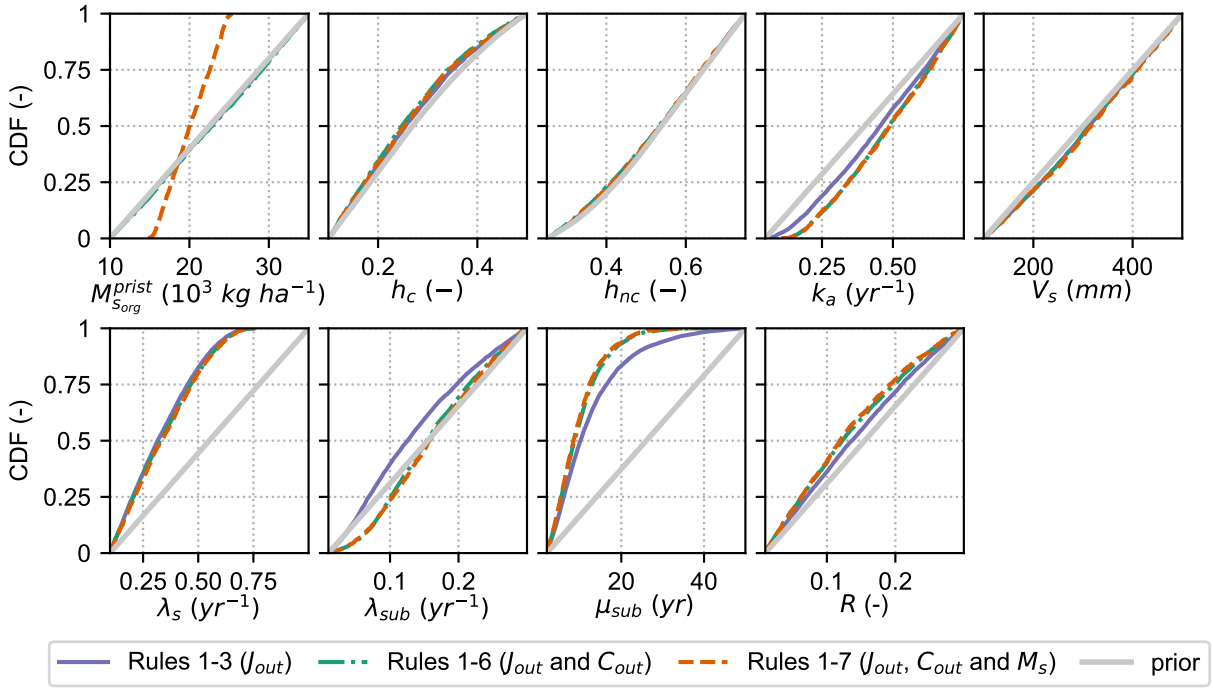


**Figure S27.** Cumulative distributions function (CDFs) of the model parameters obtained after application of Rules 1-3 (solid blue lines), Rules 1-6 (dash-dotted green lines) and Rules 1-7 (dashed red lines) and prior parameter distribution (solid grey line) for **Hemelingen**.

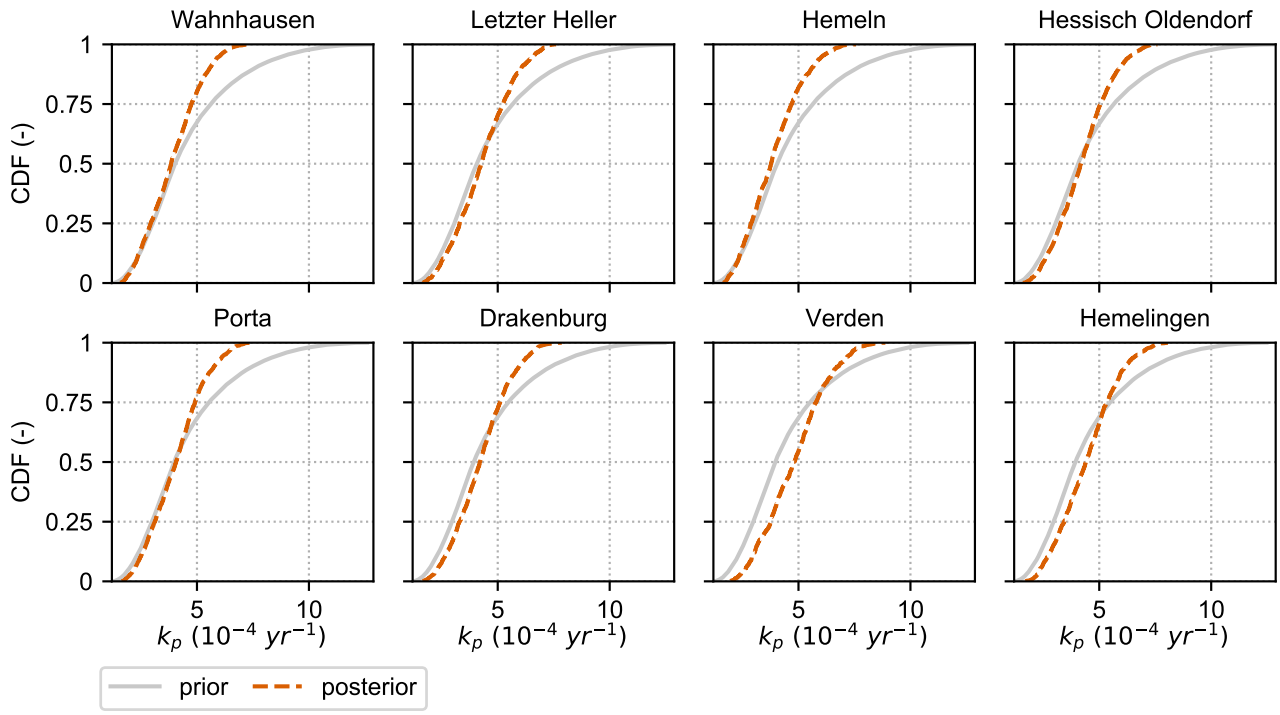


**Figure S28.** Cumulative Distribution Function (CDF) of the mineralization rate constant for the source zone organic protected N pool ($k_p$, calculated from Equation S24) in the prior simulation ensemble (before application of the soft rules) and in the posterior simulation ensemble (after application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

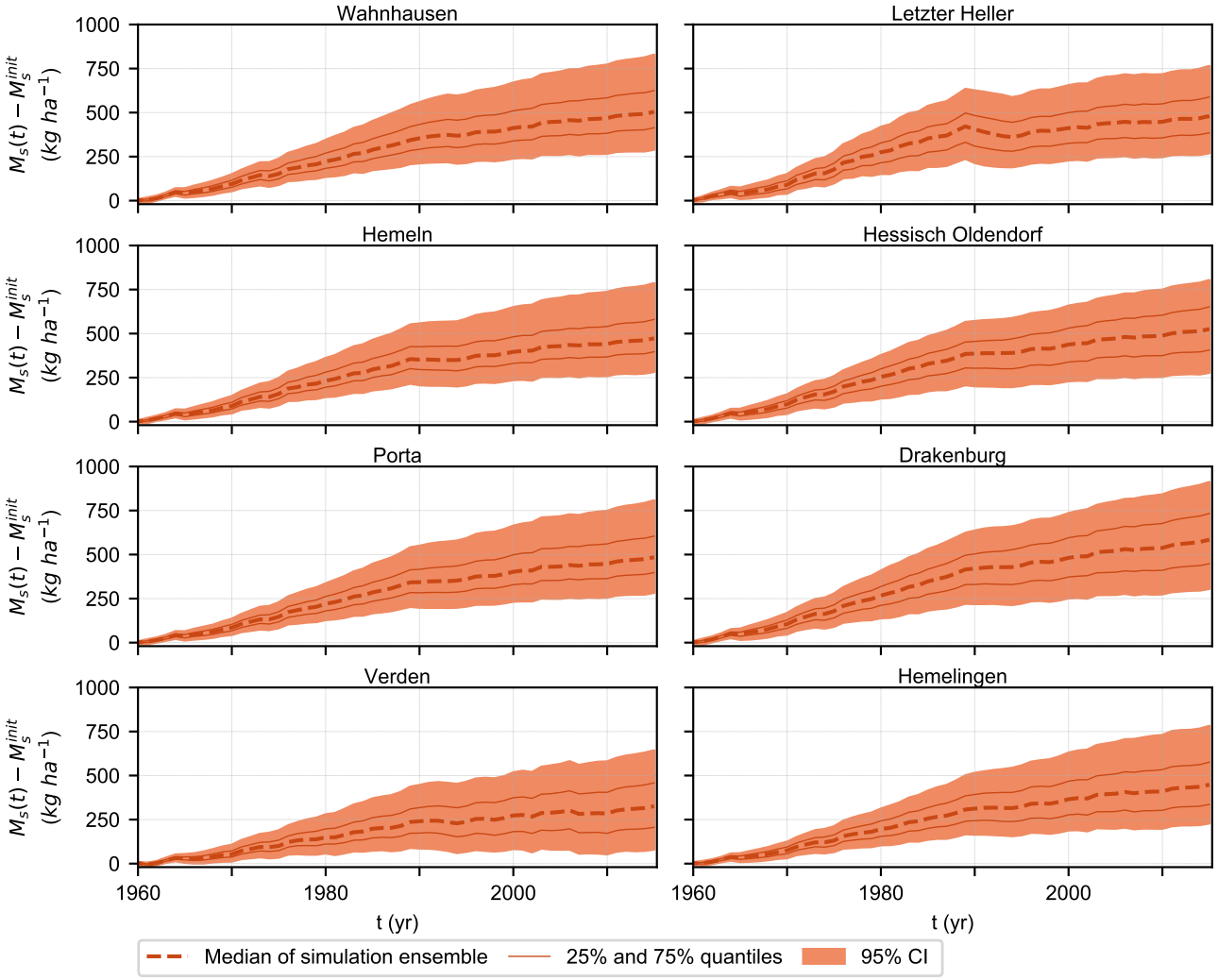


**Figure S29.** Simulated cumulative change in N storage in the source zone since 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation ensemble. Notations: $t$ is the time; $M_s$ is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

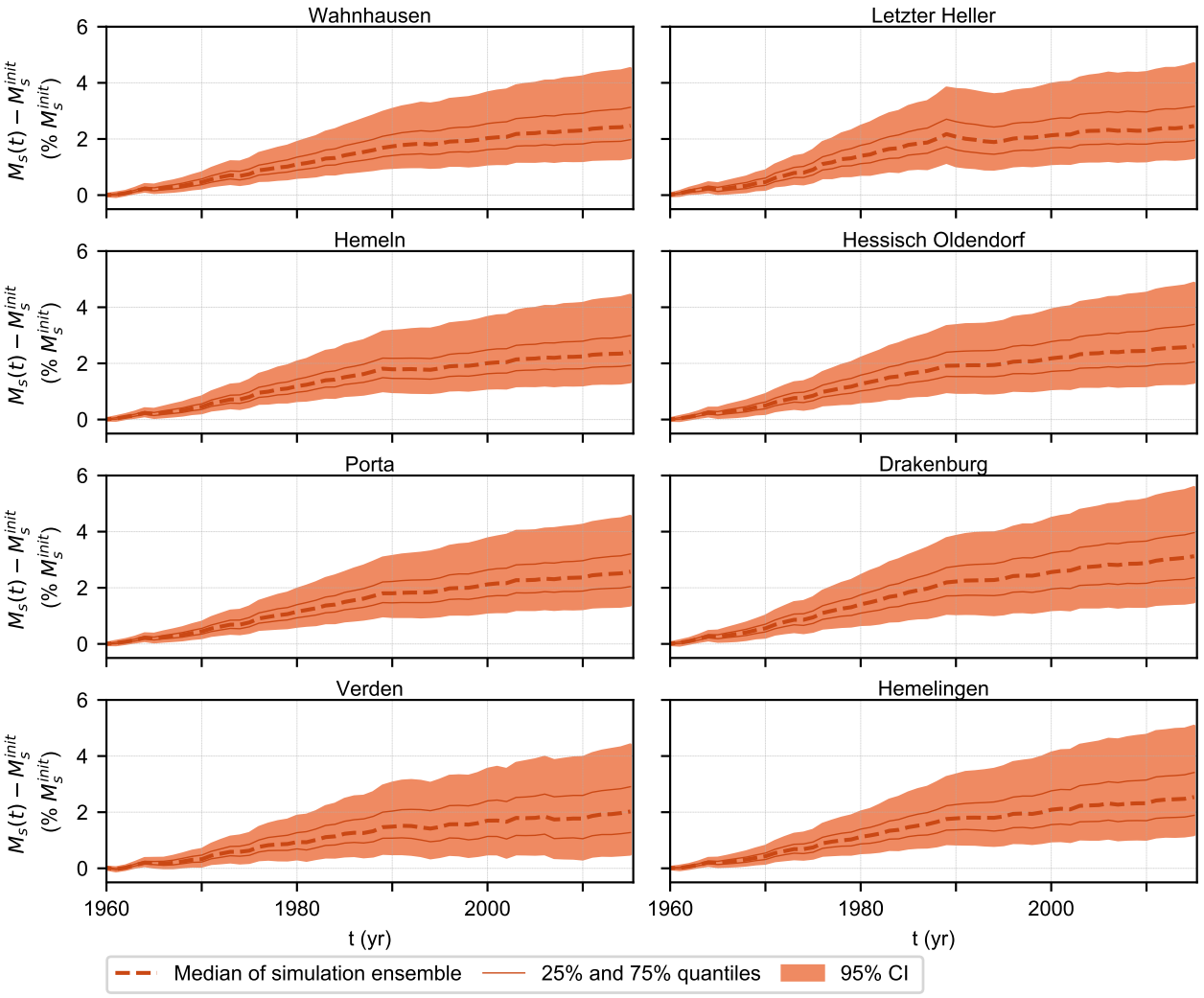


**Figure S30.** Simulated cumulative change in N storage in the source zone since 1960 as a percentage of the initial storage in 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation 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.

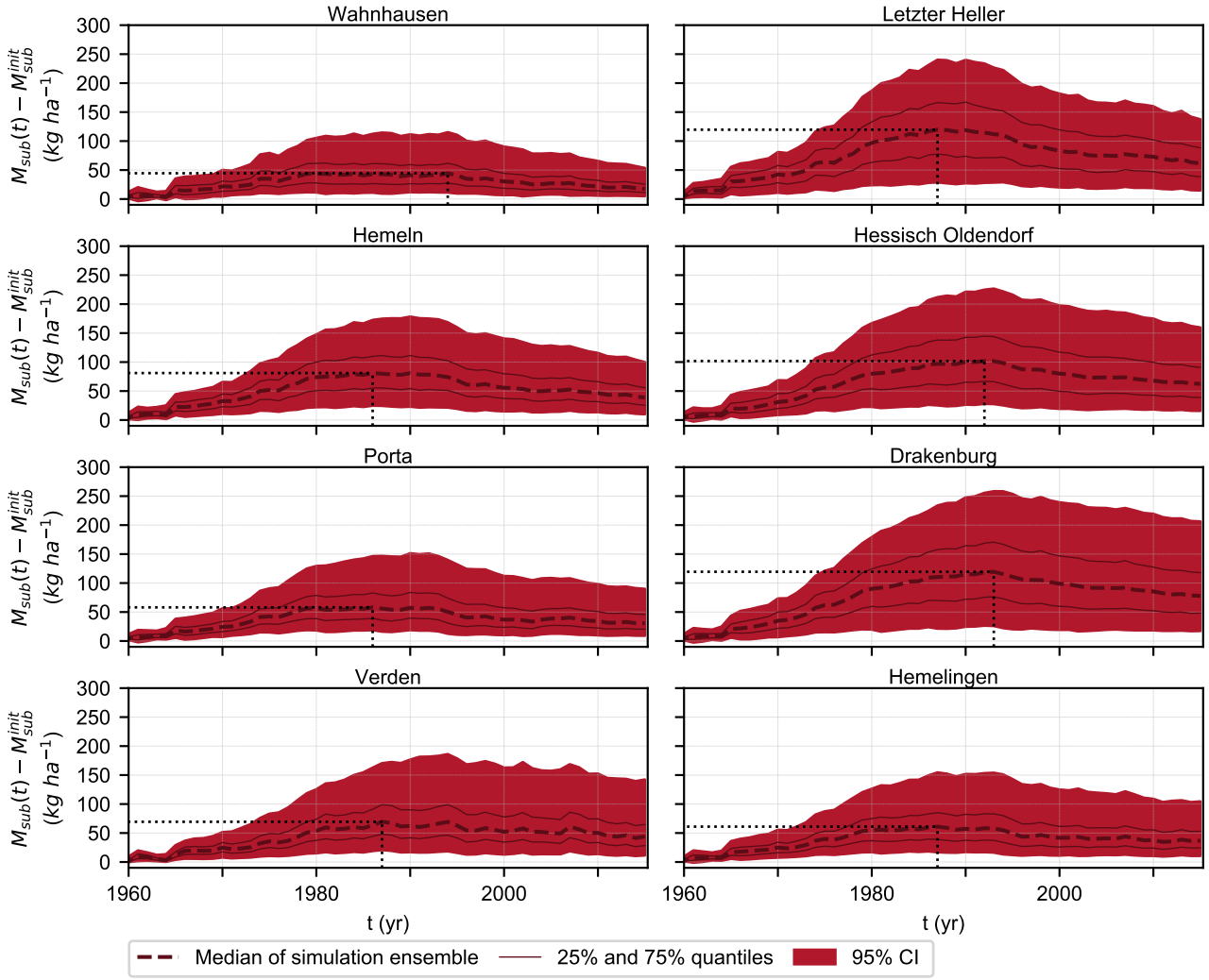


**Figure S31.** Simulated cumulative change in N storage in the subsurface since 1960. The shaded areas indicate the 95% confidence intervals, the dashed lines the median value and the solid lines the 25% and 75% quantiles in the behavioural simulation ensemble. The maximum value of the median time series 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.

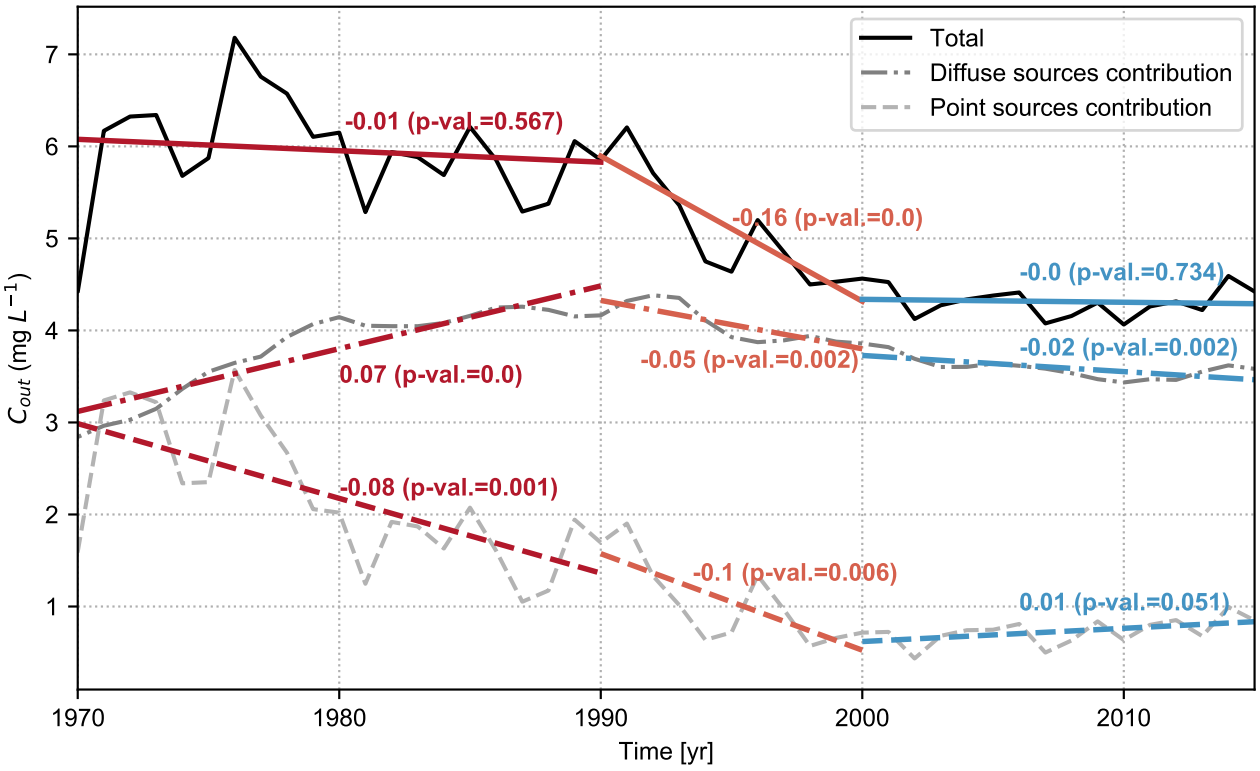


**Figure S32.** Time series of the simulated in-stream N concentration $C_{out}$ (median value of the behavioural simulation ensemble, grey lines) and linear regression lines (coloured lines) for the period 1970–2015 over the Weser River Basin (at Hemelingen). The figure reports the total N concentration, 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.

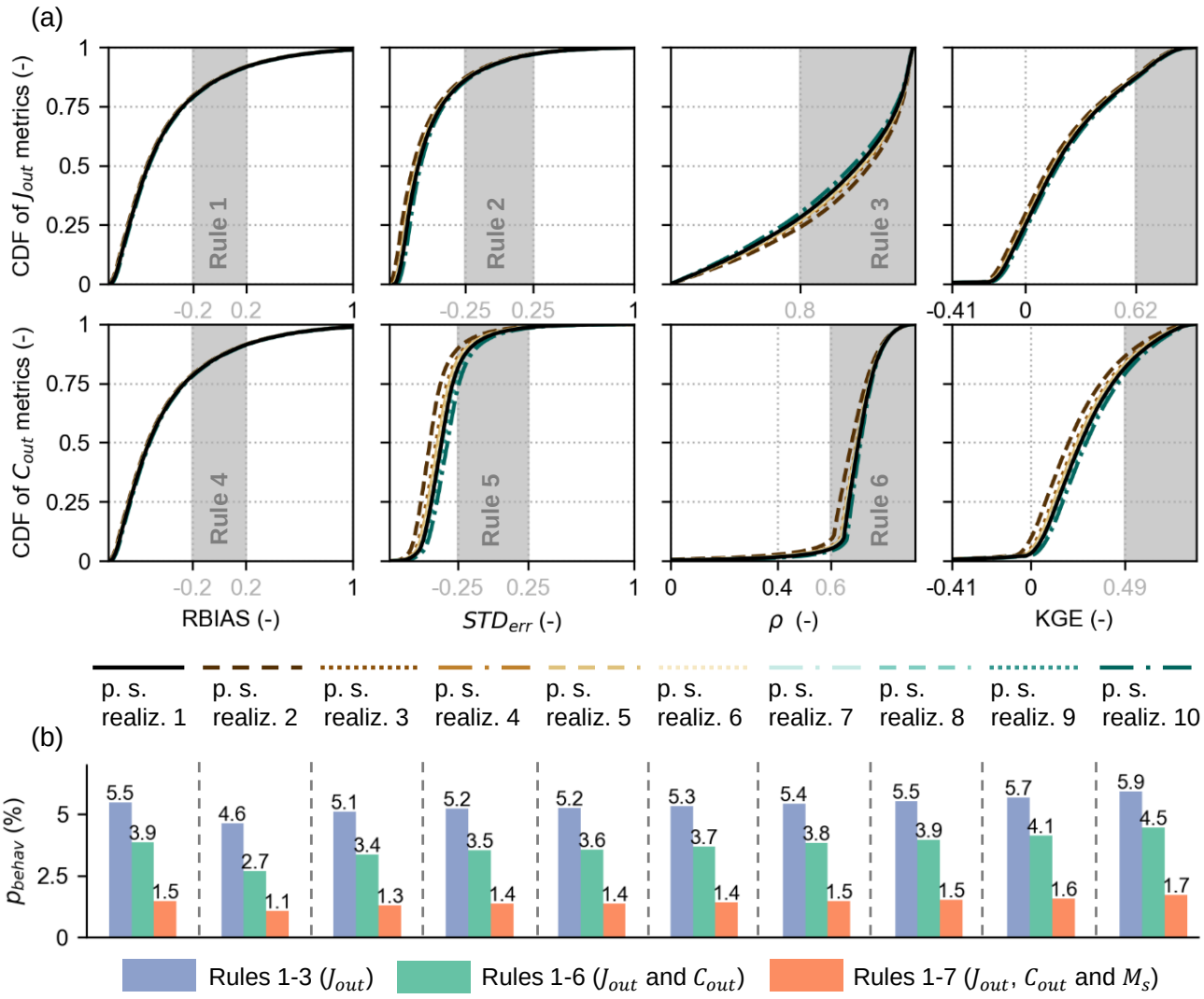


**Figure S33.** Application of the soft rules for the ten point source realizations (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading ($J_{out}$) 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 $\rho$) 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.

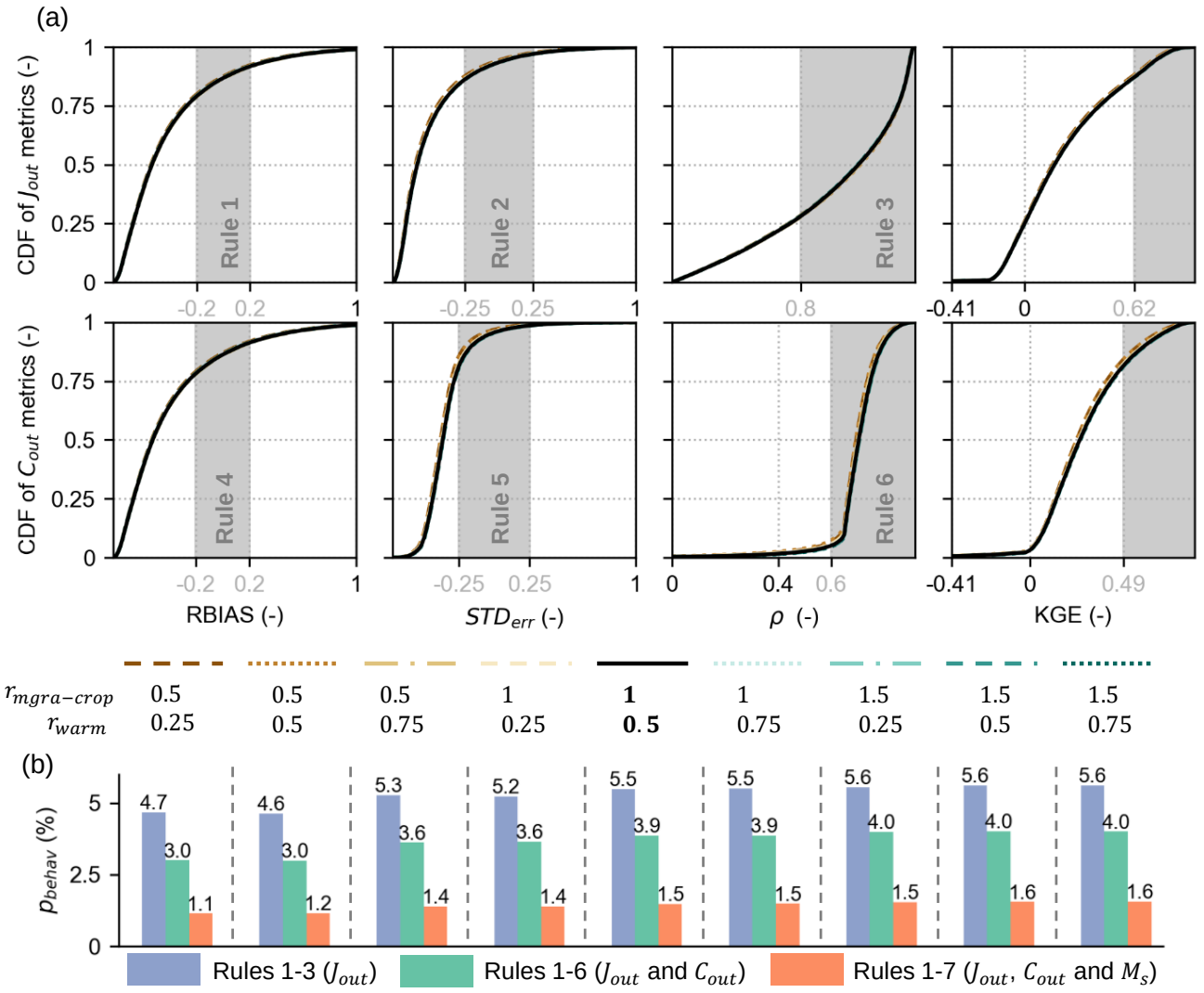


**Figure S34.** Application of the soft rules for the nine combinations of $r_{mgra-crop}$ and $r_{warm}$ (for $f_{surplus} = 1$ and the first point source realization) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading ($J_{out}$) and concentration ($C_{out}$) 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 $\rho$) 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.

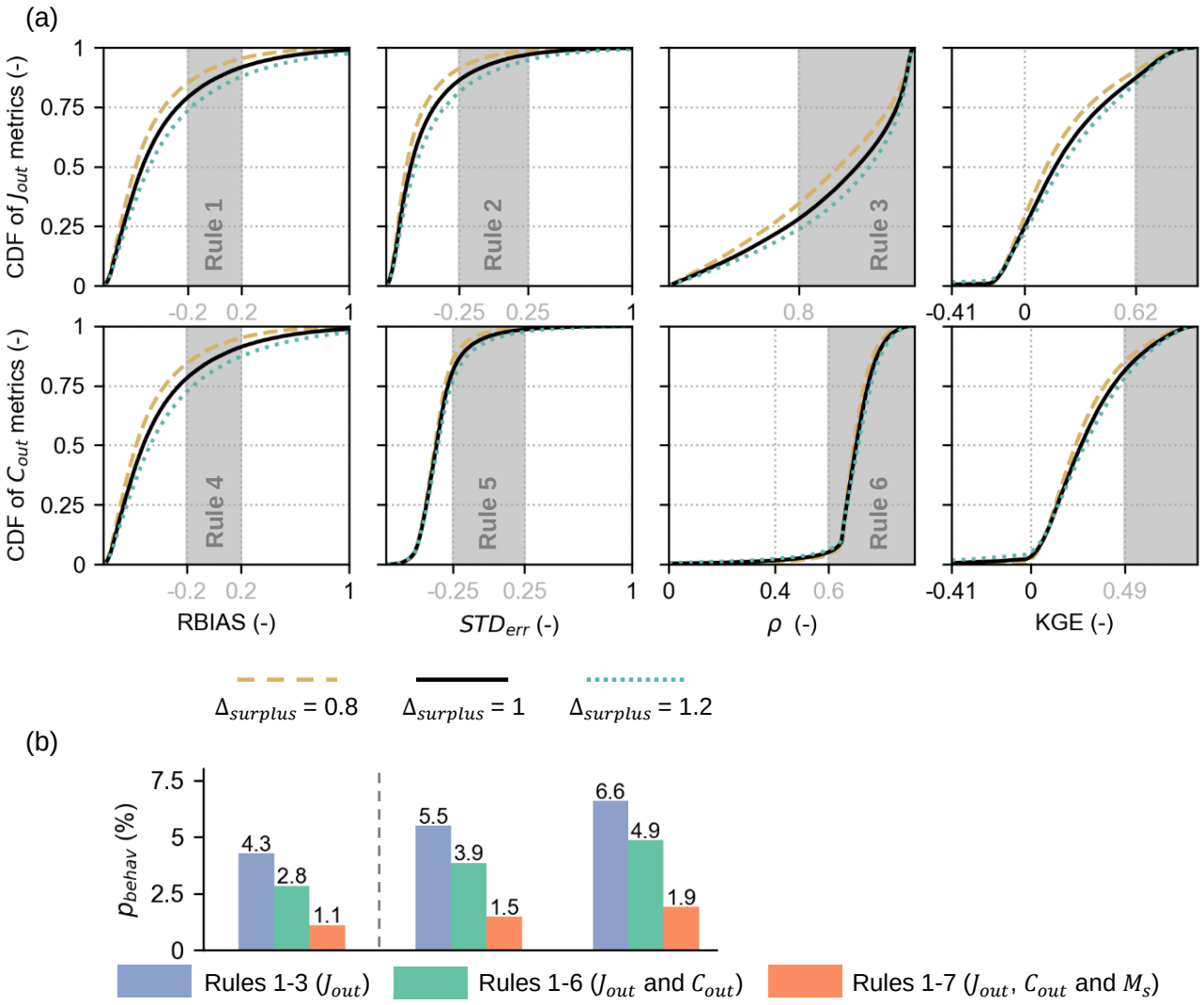


**Figure S35.** Application of the soft rules for the three realizations of $f_{surplus}$ (for $r_{mgra-crop} = 1$, $r_{warm} = 0.5$ and the first point source realization) for **Hemelingen**: (a) Cumulative Distribution Function (CDF) of the performance metrics for in-stream loading ($J_{out}$) and 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 $\rho$) 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.

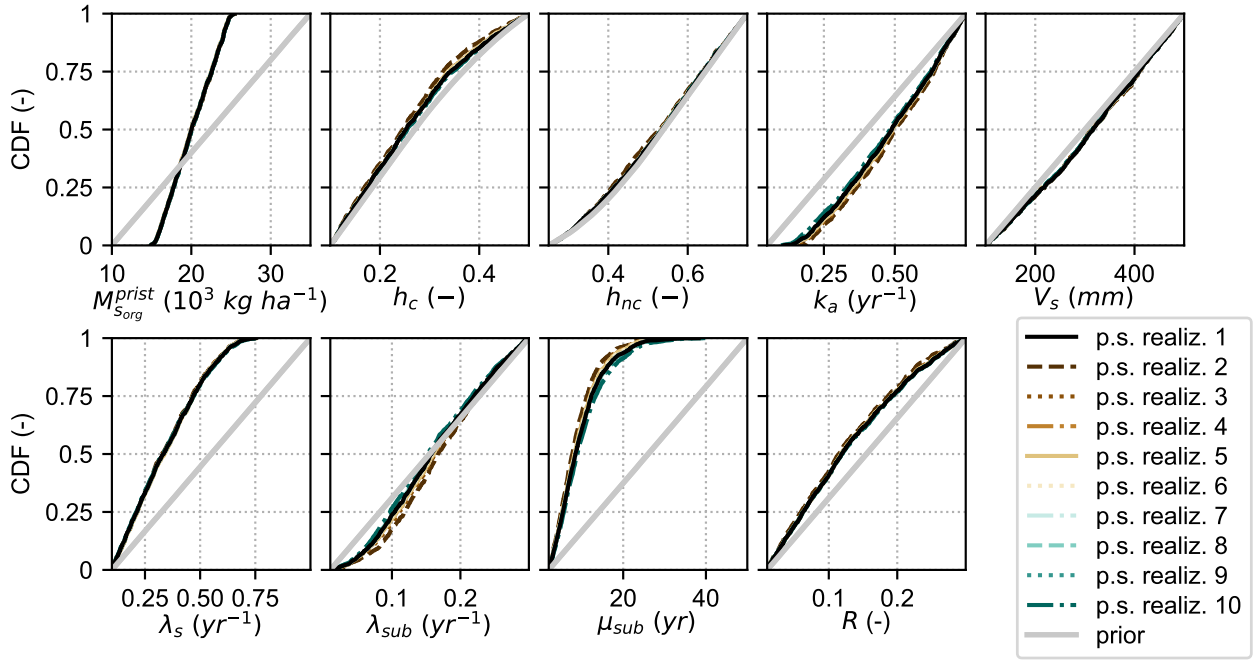


**Figure S36.** Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the ten N point sources realizations (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, $r_{mgra-crop} = 1$ and $r_{warm} = 0.5$), 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.

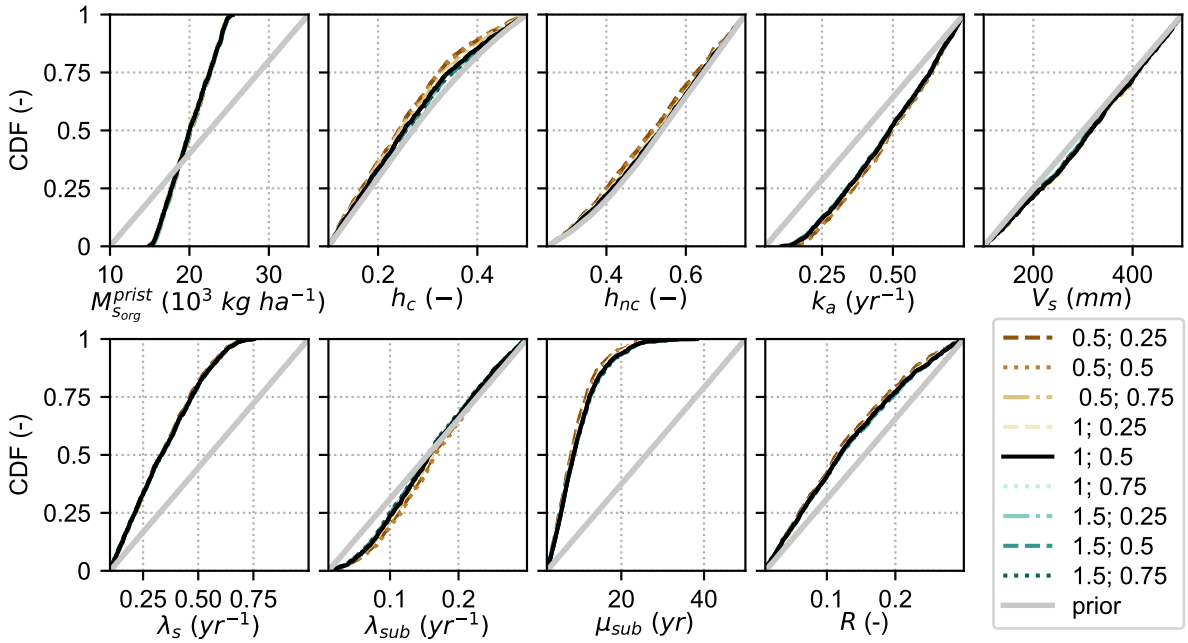


**Figure S37.** Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the nine combinations of $r_{mgra-crop}$ and $r_{warm}$ (for the baseline N surplus scenario, i.e. $f_{surplus} = 1$, and the 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.

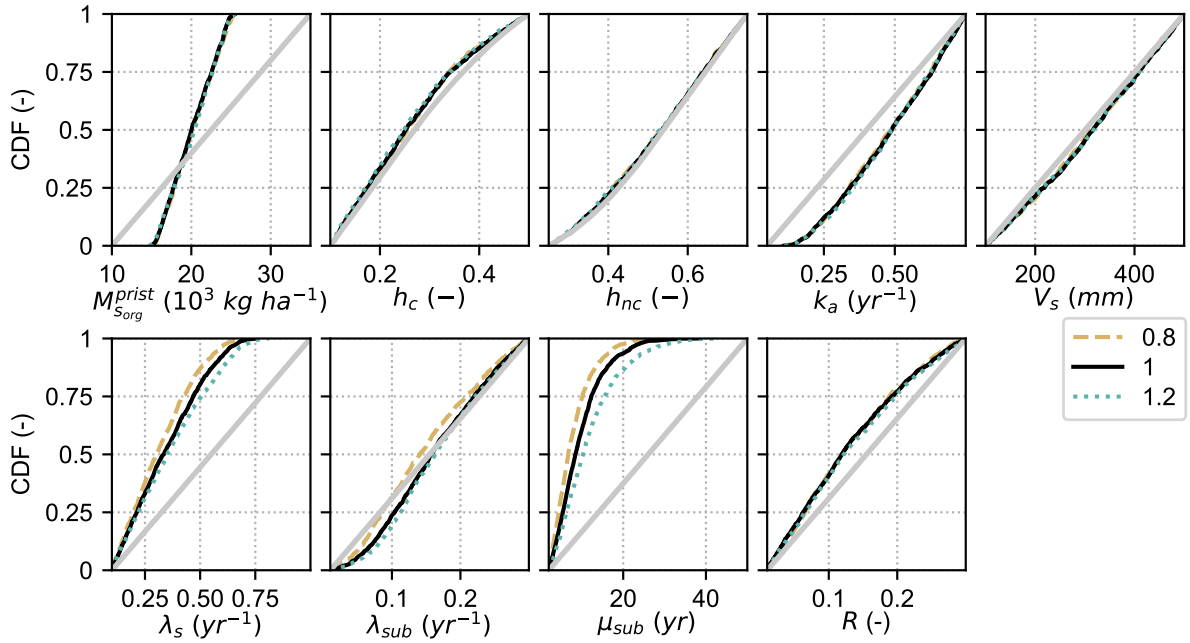


**Figure S38.** Cumulative distribution function (CDF) of the model parameters in the behavioural parameter sample, which was obtained after applications of the soft rules for the three realizations of $f_{surplus}$ (for $r_{mgra-crop} = 1$, $r_{warm}$=0.5, and the first point source realization), and 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.

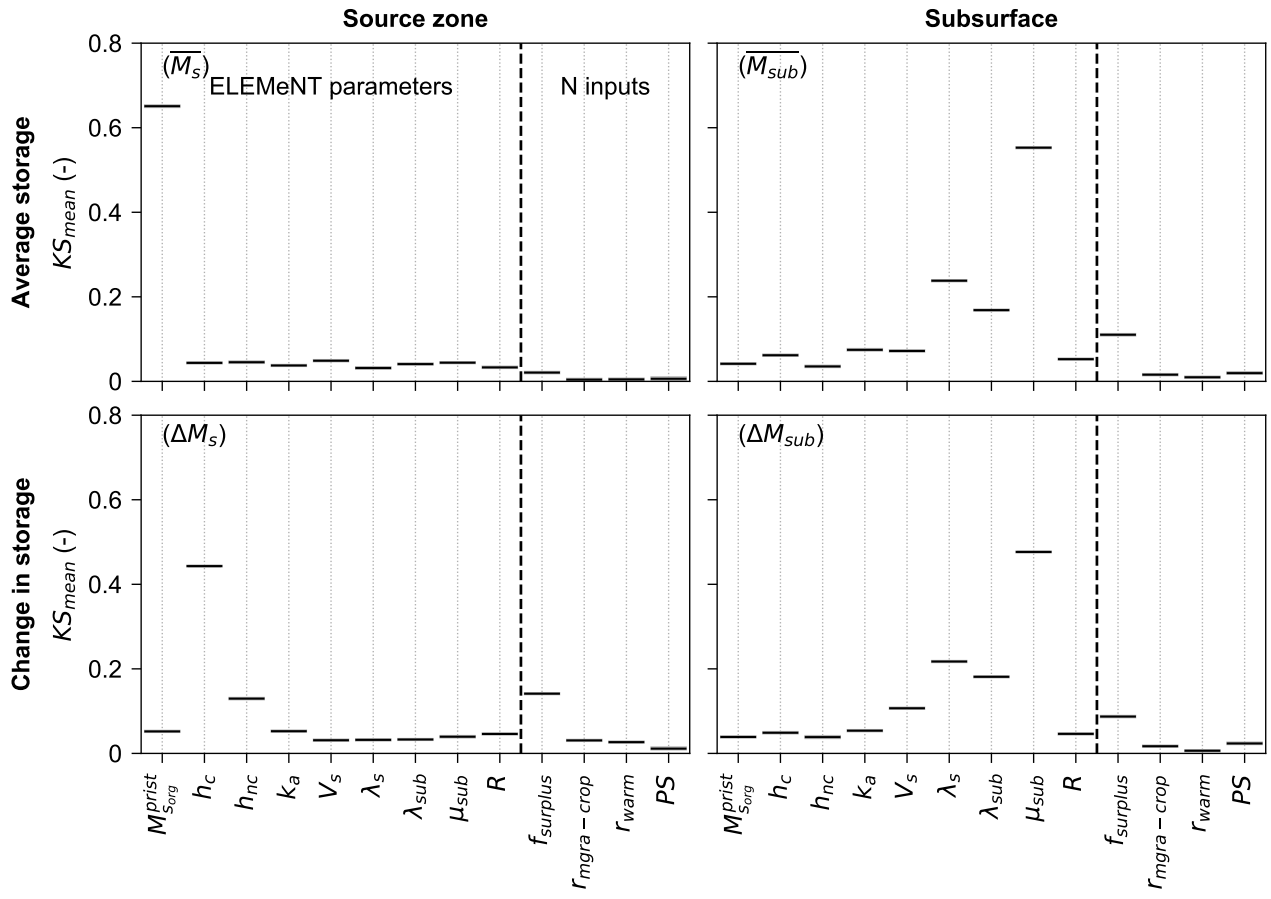


**Figure S39.** PAWN sensitivity indices ($KS_{mean}$, estimated as the mean value of KS statistic across all conditioning intervals) of the nine ELEMeNT parameters, the three parameters introduced to generate alternative N surplus realizations ($f_{surplus}$, $r_{mgra-crop}$, and $r_{warm}$), and the 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 $\Delta M_s$, and the cumulative change in subsurface N storage $\Delta M_{sub}$. 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.

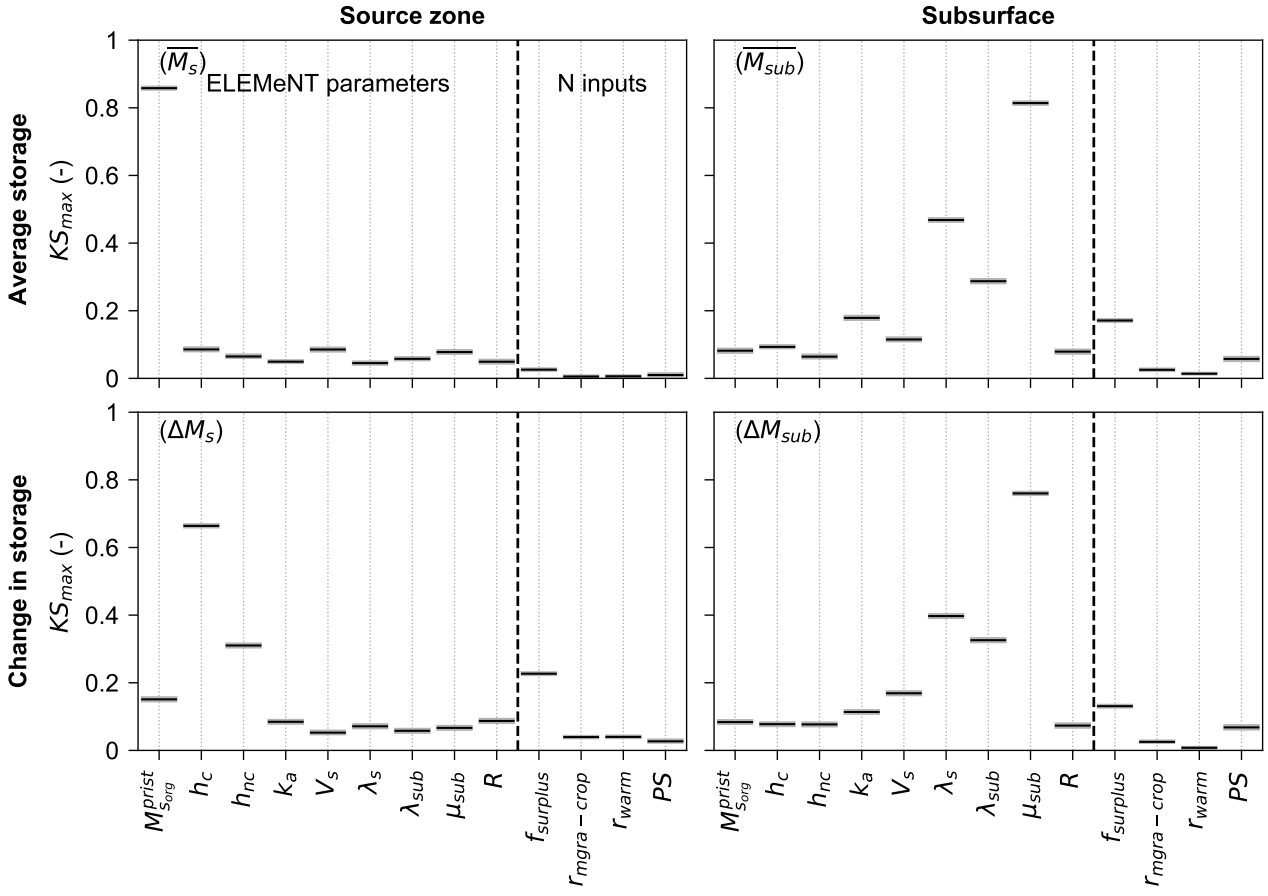


**Figure S40.** PAWN sensitivity indices ($KS_{max}$, estimated as the maximum value of KS statistic across all conditioning intervals) of the nine ELEMeNT parameters, the three parameters introduced to generate alternative N surplus realizations ($f_{surplus}$, $r_{mgra-crop}$, and $r_{warm}$), and 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 $\Delta M_s$, and the cumulative change in subsurface N storage $\Delta M_{sub}$. 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.

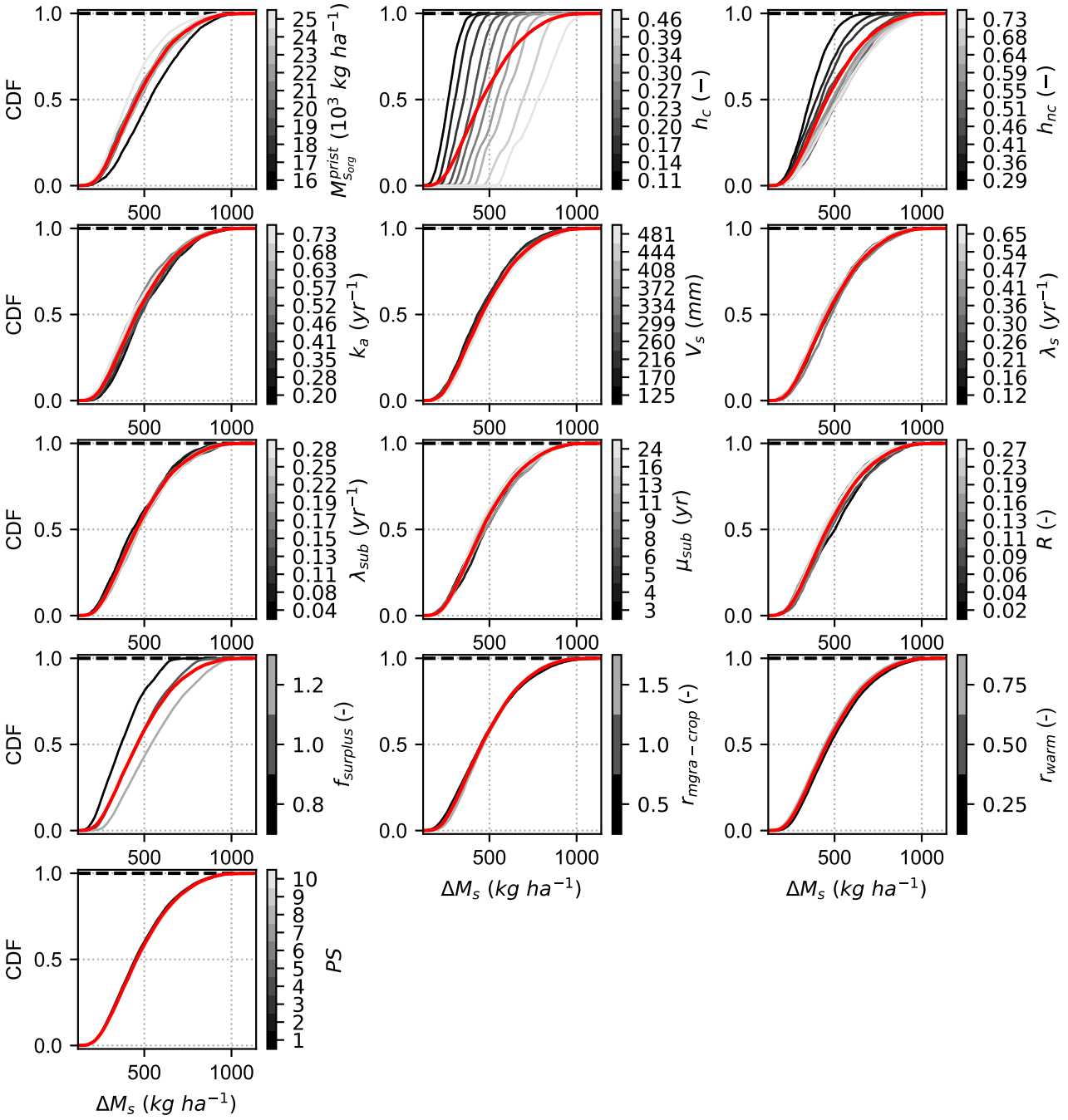


**Figure S41.** Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the change in source zone N storage over the period 1960–2015 ($\Delta M_s$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios 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.

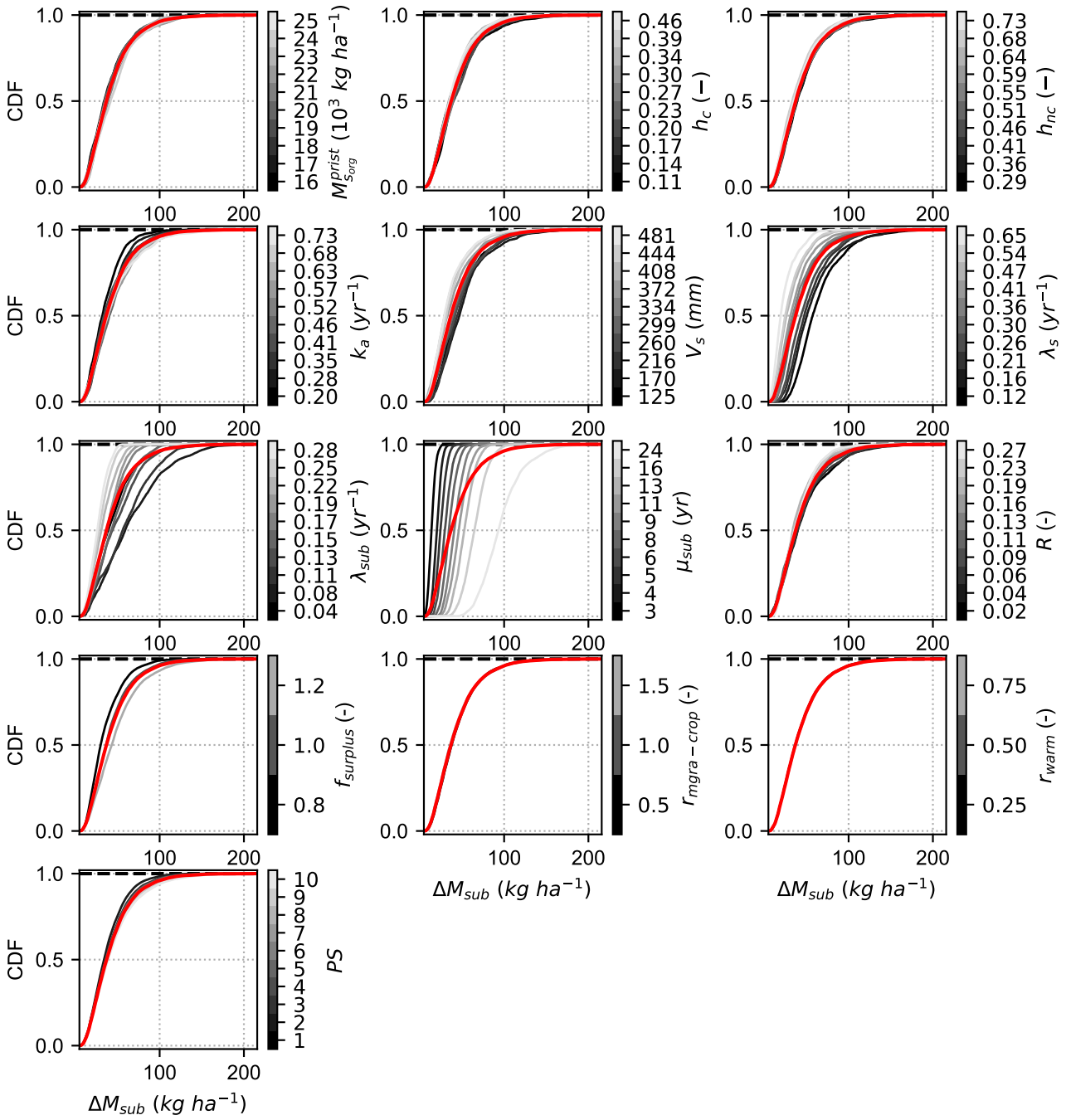


**Figure S42.** Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the change in subsurface N storage over the period 1960–2015 ($\Delta M_{sub}$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios 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.

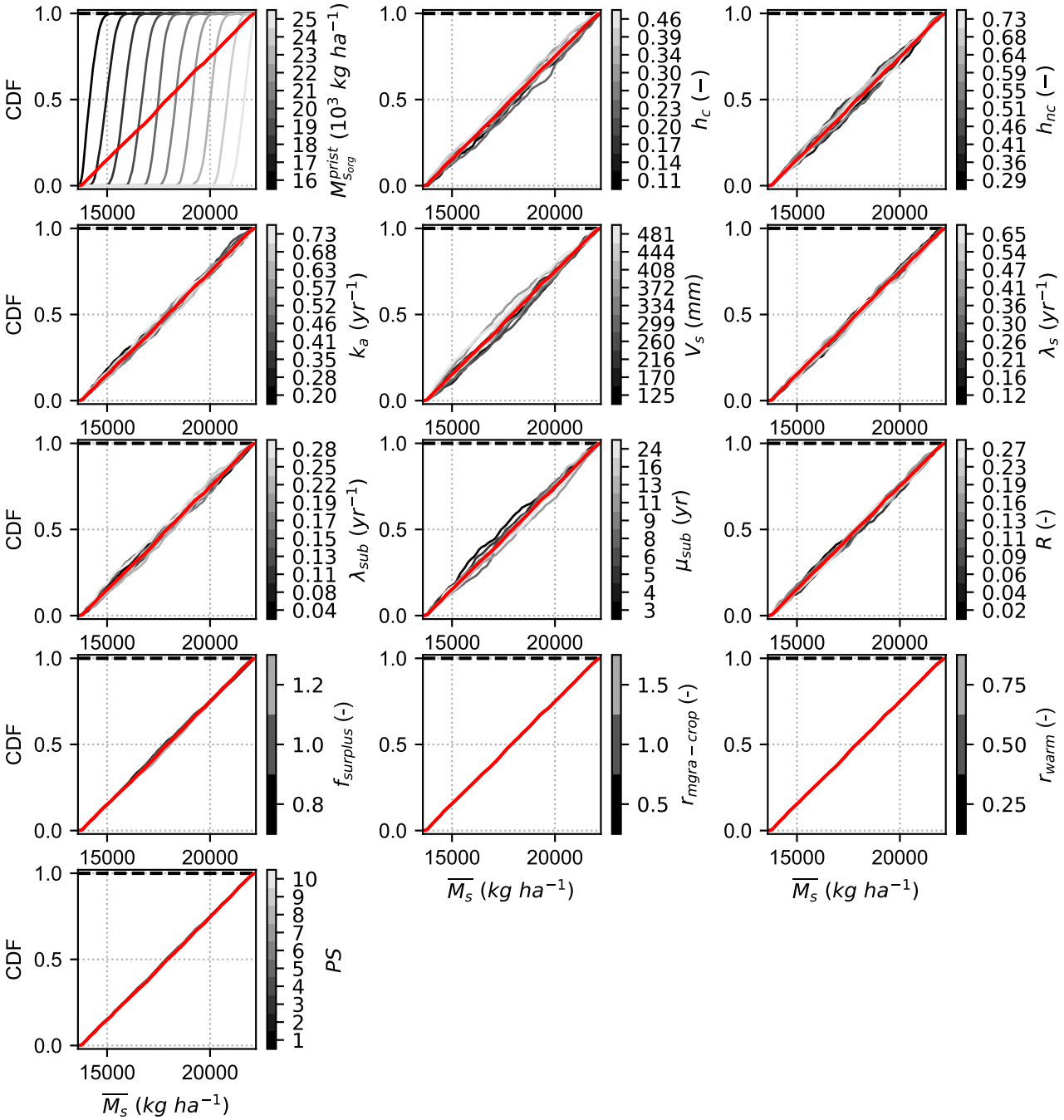


**Figure S43.** Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the average source zone N storage over the period 1960–2015 ($\overline{M_s}$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios 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.

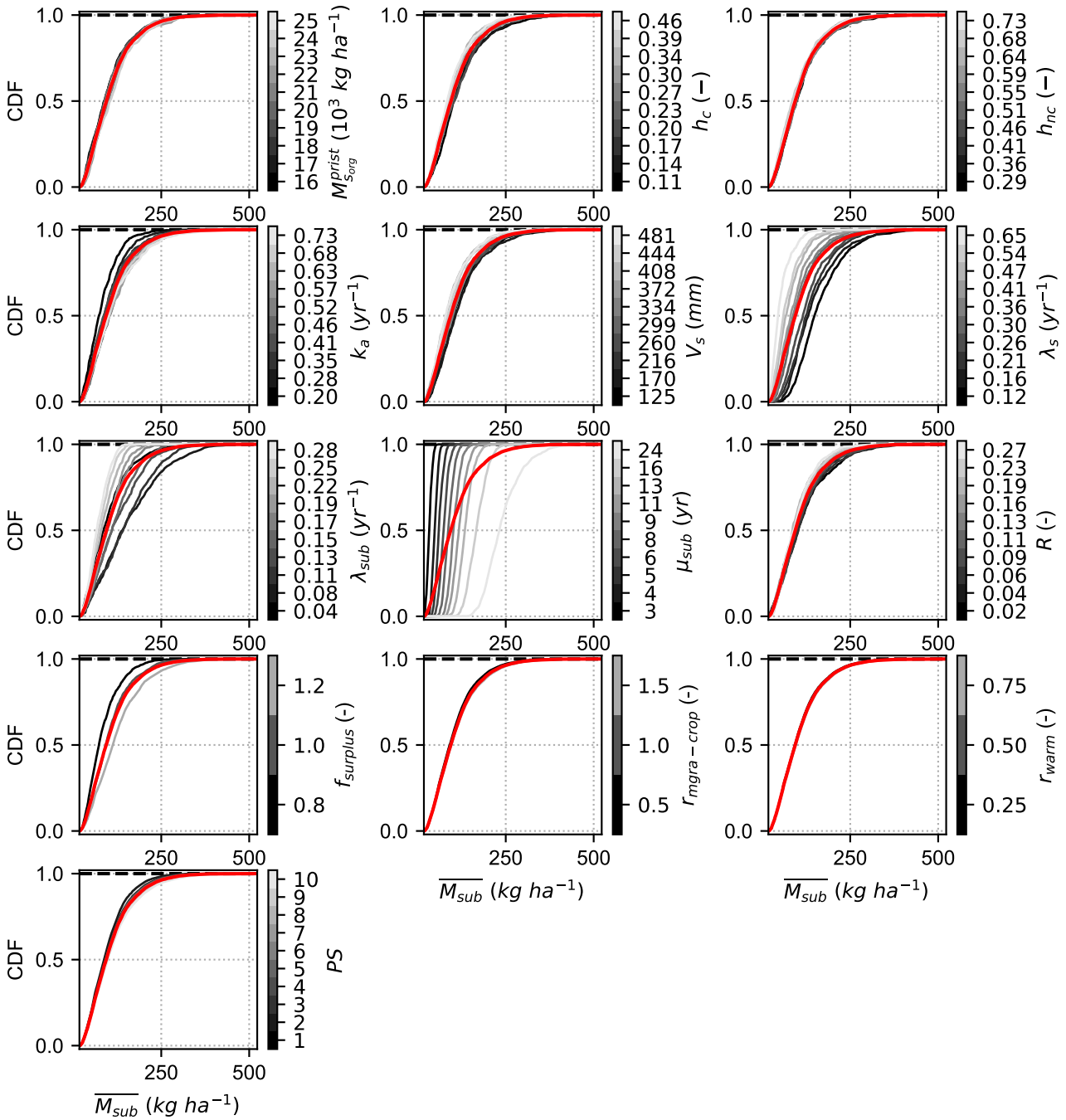


**Figure S44.** Unconditional Cumulative Distribution Functions (CDF, red line) and conditional CDFs (grey lines) of the average subsurface N storage over the period 1960–2015 ($\overline{M_{sub}}$), for the nine ELEMeNT parameters, the three parameters introduced to generate alternative N-surplus scenarios 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.