Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SCALES-Global: An annual, 0.05° global nitrogen-surplus dataset (1961–2023)

Kimberly J. Van Meter

vanmeterlab@gmail.com

The Pennsylvania State University

Nandita B. Basu

University of Waterloo

data-descriptor

Keywords:

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11515823/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SCALES-Global: An annual, 0.05° global nitrogen-surplus dataset (1961–2023)

Authors

Kimberly J. Van Meter[1,2], Nandita B. Basu[3,4,5]

Affiliations

1. Department of Geography, The Pennsylvania State University, University Park, PA, USA

2. Earth and Environmental Systems Institute, The Pennsylvania State University, University Park, PA, USA

3. Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada

4. Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, Canada

5. Water Institute, University of Waterloo, Waterloo, ON, Canada

corresponding author(s): Kimberly Van Meter (vanmeterlab@gmail.com)

Abstract

Nitrogen (N) surplus, defined as the difference between N inputs to land and N removed in harvest, is a key integrator linking food production to water-quality degradation, greenhouse-gas emissions, and biodiversity loss. Yet existing global N-surplus products are either coarse national tables, single-year snapshots, or fertilizer-only inputs, and regional gridded reconstructions commonly stop at national borders. Here, we present SCALES-Global, the first global implementation of the Scalable Computation of Anthropogenic Land Element Surpluses (SCALES) framework for consistent nitrogen accounting across subnational, national, and global scales. SCALES-Global is a bottom-up global N mass balance computed annually (1961-2023), downscaled to a 0.05° grid by mass-conserving proportional allocation from nationally reported data. The budget includes synthetic N fertilizer, biological N fixation, atmospheric deposition, livestock manure, human waste, crop N removal, and forage removal from grazing land. Surplus is calculated using three internally consistent framings, an agronomic cropland N surplus, a landscape N surplus, and a total anthropogenic N surplus, each with Monte Carlo uncertainty estimates at national and grid scales. SCALES-Global was validated against the FAO Cropland Nutrient Budget, a European soil-N-surplus reconstruction, and global gridded products, providing the continuous, fine-grained surplus trajectory required by water-quality, climate, and planetary-boundary applications.

Background & Summary

Humanity's intervention in the global nitrogen cycle is among the largest of all planetary perturbations.[1–4] Synthetic fertilizer, agricultural biological N fixation, and intensified livestock production have roughly doubled reactive-nitrogen (N) inputs to terrestrial ecosystems since pre-industrial times, driving eutrophication of fresh and coastal waters, groundwater nitrate contamination, nitrous oxide emissions, and the transgression of the proposed planetary boundary for biogeochemical flows.[5,6] The nitrogen surplus, the difference between N inputs and harvested N removal, is one of the most policy-relevant integrators of these pressures because it is directly linked to risks of nitrate leaching, surface runoff, and nitrous-oxide emissions.[7–10]

Global quantification of the nitrogen (N) surplus has developed along several complementary, but largely disconnected, lines. National budgets compiled consistently across countries provide long-term, internally consistent mass balances, but their country-scale resolution obscures substantial spatial variation within national boundaries. Global gridded datasets have substantially improved the spatial representation of individual N-cycle components, including synthetic N fertilizer[11,12] and atmospheric N deposition[13]. However, these products were developed independently, often differ in spatial and temporal resolution, and have not been harmonized into a globally consistent N surplus framework. The recently developed Historical Anthropogenic Nitrogen Inputs (HaNi) dataset[14] represents a major advance by providing annual, global, gridded estimates of most major anthropogenic N inputs, including fertilizer, manure, biological N fixation, and atmospheric deposition. However, because HaNi is an inventory of N inputs rather than a complete mass balance, it does not account for crop removal and therefore cannot directly quantify nitrogen surplus. It also does not fully resolve the human-mediated pathways that redistribute and recycle N through crop production, livestock feed, manure generation and reuse, and human consumption and waste.

Parallel to these global efforts, high-resolution annually resolved, gridded N surplus datasets have been developed for specific regions e.g. for Europe,[15] China,[16] and the United States (gTREND-N).[17] These regional products demonstrate that spatially explicit surplus reconstruction is feasible, but their geographic coverage, methods, and underlying data remain region-specific. Consequently, no existing global dataset provides an annually resolved, fine-resolution (~5 km) nitrogen surplus that integrates harmonized N inputs and crop removal, reconciles gridded estimates with national nutrient budgets, represents N transfers and recycling across crop, livestock, and human systems, quantifies uncertainty, and is independently evaluated against national and regional benchmarks.

A further limitation of existing nitrogen mass-balance studies is that human nitrogen fluxes are rarely incorporated into budgets at either the country or global scale. Where human waste is represented at all, it is typically entered as a single lumped anthropogenic input as total N excretion or a population-scaled term, without resolving its fate through sanitation and wastewater treatment.[18] While a separate body of work has modeled global wastewater nitrogen and its removal during treatment,[19,20], these efforts were developed for water-quality assessment and have not been integrated into terrestrial nitrogen-surplus accounting. Accordingly, estimated N surplus magnitudes in existing datasets either grossly overestimate N inputs by not taking into account wastewater treatment, or underestimate N inputs by not including human N at all.

To address these gaps, we have developed SCALES-Global, the first global implementation of the SCALES framework (Scalable Computation of Anthropogenic Land Element Surpluses). SCALES provides a unified, scale-consistent approach to nutrient-surplus accounting by applying consistent budget definitions and coefficient lineages across spatial scales, from country to grid cell. Here, annual national

N budgets are constructed for the years 1961-2023 from FAOSTAT activity data, complementary environmental datasets, and published process-based and empirical coefficients (**Figure 1**). Among these components, SCALES-Global explicitly resolves the human nitrogen cycle, partitioning excreted nitrogen across pathways that include recycling to cropland, removal during wastewater treatment, and discharge to surface waters, so that human fluxes are carried within the same mass-balance framework as agricultural inputs. The current budgeting approach can reproduce the national-scale FAOSTAT Cropland Nutrient Budget (with which it shares most input data and coefficients), and the final products are also corroborated against independent datasets. The national budgets are subsequently downscaled to a 0.05° grid using process-specific spatial proxies and a mass-conserving allocation procedure that preserves national total while resolving spatial heterogeneity. The resulting dataset provides annual, gridded estimates of nitrogen surplus magnitudes, individual components of the mass budget, and associated uncertainties.

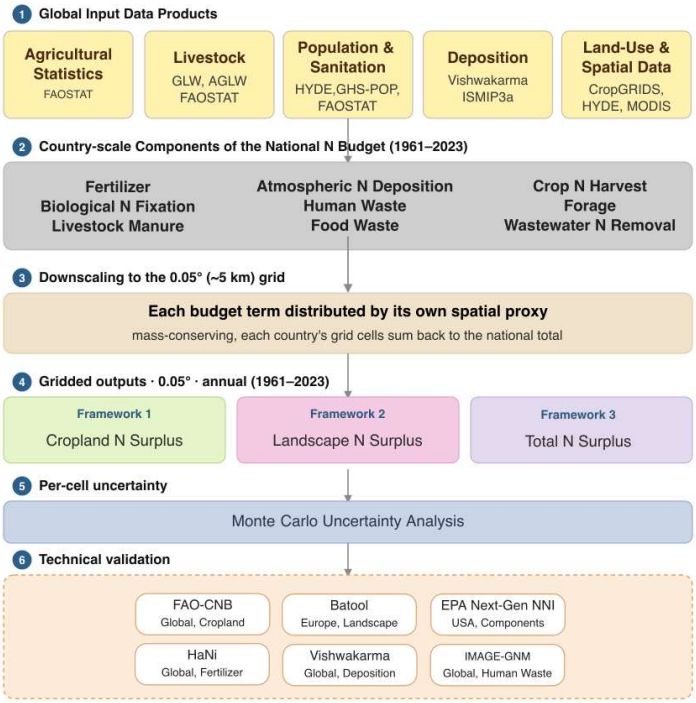


Figure 1. Workflow for constructing the SCALES-Global nitrogen surplus dataset. Public datasets are integrated to produce annual (1961–2023), 0.05° (~5 km) gridded nitrogen (N) surplus estimates, with associated uncertainty. National N budgets are first computed from FAOSTAT activity data, other relevant environmental datasets, and appropriate coefficients to calculate N inputs to the landscape using a soil-surface ("as-applied") mass balance. Each budget component is then independently downscaled to the grid using process-specific spatial proxies while conserving national totals. The released products comprise three gridded surplus frameworks: cropland surplus, landscape surplus, and total anthropogenic surplus. Per-cell uncertainties are estimated by propagating the uncertainty of each budget term from a national Monte Carlo analysis to the grid. Technical validation includes consistency checks against the FAO Cropland Nutrient Budget as well as independent comparisons with published regional and global nitrogen budget datasets.

SCALES-Global provides three internally consistent accounting frameworks derived from a common underlying nitrogen budget. The first is a cropland surplus, which quantifies the nitrogen surplus across

all cropland within each 5-km grid cell. It includes cropland-specific nitrogen inputs from synthetic fertilizer, manure, biological nitrogen fixation, atmospheric deposition, and recycled human nitrogen applied to cropland, and subtracts nitrogen removed through crop harvest. It follows the same input-output framework as the FAO Cropland Nutrient Budget, but it additionally includes recycled human nitrogen applied to cropland. The second framework is a landscape surplus, which includes all components of the cropland surplus together with nitrogen inputs and forage removals associated with pasture- and rangeland-based agriculture, as well as recycled human waste on non-agricultural land. The third is a total anthropogenic surplus (**Figure 2**), which includes all components of the landscape surplus and additionally accounts for human nitrogen released to the environment through wastewater treatment plant outflows. A single budgeting approach therefore supports research communities that have historically relied on different and often incompatible nitrogen accounting approaches. In addition, Monte Carlo uncertainty propagation provides uncertainty estimates for both the national budgets and the resulting gridded products, allowing users to evaluate the confidence associated with individual budget terms and nitrogen surplus estimates.

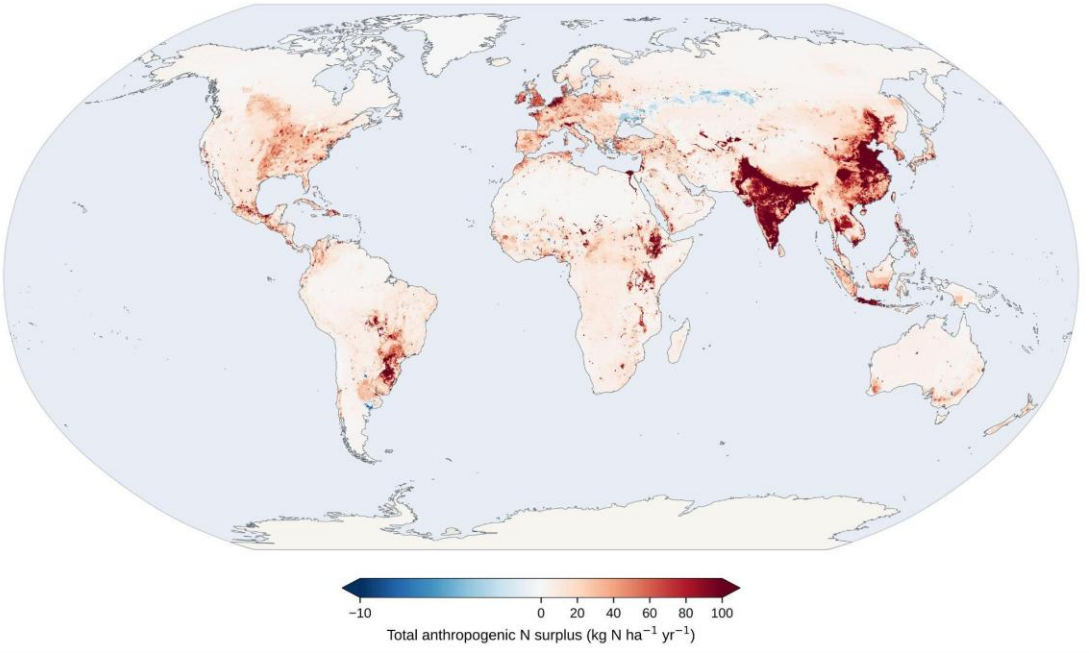


Figure 2. Global total anthropogenic nitrogen surplus in 2023. Annual total anthropogenic nitrogen (N) surplus at 0.05° (~5 km) spatial resolution. This accounting view extends conventional cropland nutrient surplus accounting by incorporating grazing-land nitrogen flows and wastewater-derived nitrogen discharged to the environment, thereby representing anthropogenic nitrogen surplus across the entire terrestrial landscape. Values are expressed as kg-N ha^{-1} yr^{-1}.

Methods

SCALES-Global provides annual (1961-2023), 0.05° (~5.5 km; 7200 × 3600 cells) estimates of components of the terrestrial nitrogen (N) budget and the resulting nitrogen surplus magnitudes . The released products include annual gridded estimates of synthetic fertilizer application, livestock manure, biological N fixation, atmospheric N deposition, recycled human N, crop N removal, and ammonia (NH_3) volatilization losses from manure and synthetic fertilizer, together with derived cropland, landscape, and

total anthropogenic N surplus estimates. All gridded surplus products are reported in kg-N ha^{-1} yr^{-1} (component fluxes in kg-N $cell^{-1}$ yr^{-1}) and are accompanied by uncertainty estimates derived from Monte Carlo propagation.

In SCALES-Global, each national budget term is built bottom-up from activity data and coefficients, then spatially disaggregated top-down to the grid. National N budgets are first computed directly from publicly available FAOSTAT activity data, complementary environmental datasets, and literature-derived coefficients. Each budget component is then independently downscaled to a 0.05° grid using a process-specific spatial proxy while conserving national totals, such that the sum of all grid cells within each country reproduces the corresponding national budget term. This approach provides spatially explicit estimates while maintaining consistency with the underlying national accounting framework. The one exception to this approach is atmospheric deposition; its non-cropland component is taken directly from a gridded product, so for this term the grid is primary and the national value is instead recovered by aggregation.

Spatial downscaling framework

Every term of the nitrogen budget, with the exception of atmospheric deposition, is first estimated at the national scale from FAOSTAT activity data and the appropriate coefficients, and then is distributed to the SCALES-global 0.05° grid using a dasymetric mapping approach. Dasymetric maps display statistical data in meaningful spatial zones rather than at the original administrative scales (countries, counties, etc.) at which the data is originally collected.[21,22] Here, each national value $X(i,t)$ is distributed to the grid cells m of country i in proportion to a term-specific spatial proxy, $\varphi_X(m,t)$. This proxy represents the within-country spatial distribution of the corresponding flux and is derived from existing spatial datasets or appropriate proxies such as cropland area, livestock density, or population (**Table 1**). For each country and year the proxy is normalized so that its dasymetric weights, $w_X(i, m, t)$, sum to one within the country:

$$w_X(i,m,t) = \frac{\phi_X(m,t)}{\sum_{m' \in i} \ (\phi_X(m',t))} \quad \text{(Eq. 1)}$$

where m' is the summation index over all grid cells in country i.

The national value is then allocated to grid cells as:

$$X(i,m,t) = X(i,t) * w_X(i,m,t) \quad \text{(Eq. 2)}$$

Because the weights sum to one within each country, the gridded values conserve the national budget term exactly, i.e. the sum of all grid cells within a country reproduces its national value. This mass-conserving allocation is applied independently to each component before the cropland, landscape, and total anthropogenic surpluses are formed.

As described here, two land bases underpin all of the agricultural terms. The cropland area of each cell in each year is taken from HYDE 3.3,[23] which reconstructs annual cropland extent, and the division of that cropland across specific crops is taken from CROPGRIDS,[24] which maps harvested and physical area for 173 crops at 0.05° for the year 2020. Crop-specific terms, e.g. N fertilizer, crop N removal, biological N fixation, are placed based on the harvested area of the relevant crop in year t, so that each flux lands only where that crop is grown; terms that are not crop-specific, e.g. cropland-applied manure, are placed based on the total cropland area.

Grazing land is taken from Global Pasture Watch (GPW),[25] which maps grassland extent at 30-m annually (2000-2022) and distinguishes two different classes: (1) cultivated grassland (intensively managed pasture), and (2) natural/semi-natural grassland (rangeland). We aggregate the 30-m classes to the SCALES-Global 0.05° grid as per-cell areas and form two bases: a narrow managed-pasture base, used for distribution of grassland fertilizer, and a broader grazing base (cultivated + rangeland), used to downscale pasture-deposited manure, forage removal, and pasture biological N fixation. Each grid year uses the Global Pasture Watch map for that year. Prior to 2000, the map is clamped to the year 2000, and for 2023, we retain the 2022 map.

Note that satellite grassland products, Global Pasture Watch included, have documented limitations in classifying intensively managed pasture, including underprediction of grassland extent and confusion between managed grassland and cropland.[25,26] These limitations create two distinct issues for the grazing-land terms, which we address separately. First, in a small number of heavily farmed countries such as the Netherlands, Slovenia, Finland, Japan), the mapped managed-pasture area is implausibly small relative to the reported permanent pasture, so that a large grassland fertilizer input would be concentrated onto a very small detected area. To prevent such implausible N loadings, we cap the combined fertilizer N+manure N magnitudes at a physically defensible maximum of 350 kg-N ha^{-1} yr^{-1}.[27,28] Where a cell would exceed this threshold, the excess is re-allocated in a cascade that conserves the national total, from GPW managed pasture to GPW rangeland and, where the GPW rangeland is too small to absorb the input at a plausible rate, to the MODIS MCD12C1-predicted grassland extent.[29,30]

Second, and separately, FAOSTAT quantifies pasture-deposited manure from livestock production-system assumptions rather than from mapped land cover. In dense-livestock, low-pasture systems, i.e. parts of South and Southeast Asia, this assigns a large pasture-manure nitrogen flux to countries with little or no mapped grazing land, where animals actually graze crop residues or are stall-fed and their manure returns to cropland. The result, during the spatial downscaling process, is pasture-deposited manure with little or no pasture to receive it (Pakistan 3.1 Tg, Indonesia 1.5 Tg, Bangladesh 0.8 Tg, all with near-zero mapped grazing land). Here, we reconcile this mismatch at the national scale by reassigning the excess pasture manure to cropland (see the Livestock Manure N section below). With that reallocation applied, the grazing-nitrogen ceiling and its cascade act primarily as a residual safeguard against fertilizer over-concentration in the few countries where managed pasture is under-mapped.

Where a source dataset is coarser than the model grid, e.g. the Zhang manure dataset (5 arcmin)[31] and the Coello fertilizer dataset (5 arcmin)[12], the coarser values are redistributed to the finer land base by area-conservative overlap, which conserves the mass of each coarse cell exactly while resolving the sub-cell pattern from the finer land map (see the fertilizer and cropland manure sections below).

Gridded surpluses are then obtained by summing the relevant component masses within each cell and dividing by the appropriate area, i.e. cropland area for the cropland surplus, total land area for the landscape and total surpluses. The downscaling locates each national flux within grid cells according to the best available proxy but does not resolve sub-cell variability.

Table 1. Downscaling proxies. For each budget term X, the within-country spatial weight is built from the proxy ϕ_X below and used to distribute the national values (Eqs. 1-2).

Budget term	Symbol	ϕ_X	Data sources
Synthetic fertilizer, cropland	$FERT_{crop,net}$	$\phi_{FERT_{crop}}(m,t) = \sum_j FERT_{crop}(j,m,t)$	Coello (2025); CROPGRIDS[12,24]; HYDE 3.3[23]
Synthetic fertilizer, managed pasture	$FERT_{mpast,net}$	$\phi_{FERT_{mpast}}(m,t) = AREA_{mpast}(m,t)$	Global Pasture Watch[25]
Manure applied to cropland	$MAN_{crop,net}$	$\phi_{MAN_{crop}}(m,t) = MAN_{crop}(m,t)$	Zhang et al. (2017); CROPGRIDS[24,31]; HYDE 3.3[23]
Manure deposited on pasture	$MAN_{graze,net}$	$\phi_{MAN_{graze}}(m,t) = M_{man,graze}(y,t) * \frac{AREA_{graze}(m)}{\sum_{m' \in y} AREA_{graze}(m')}$	Zhang et al.; Global Pasture Watch[25,31]
Biological N fixation: grain legumes	$BNF_{crop,leg}$	$\phi_{BNF_{crop,leg}}(j) = AREA_{harv}(m,j)$	CROPGRIDS[24]; HYDE 3.3[23]; Global Pasture Watch[25]
free-living	$BNF_{crop,free}$	$\phi_{BNF_{crop,free}}(j) = AREA_{harv}(m,j)$	
pasture	BNF_{graze}	$\phi_{BNF_{graze}} = AREA_{graze}(m,t)$	
Human Waste N	$FOOD_{waste}$ $HUM_{terr,nc}$ $HUM_{landfill}$	$\phi_{FOOD_{waste}} = \phi_{HUM_{terr,nc}} = \phi_{HUM_{terr,nc}} = POP(m,t)$	GHS-POP[32]
	$HUM_{effluent}$ $HUM_{c,untreat}$	$\phi_{HUM_{effluent}} = \phi_{HUM_{effluent}} = POP(m,t) * urb(m,t)$	GHS-POP[32]
	$HUM_{crop,nc}$	$\phi_{HUM_{crop,nc}} = AREA_{crop}(m)$	HYDE 3.3[23]
	$HUM_{crop,bio}$	$\phi_{HUM_{crop,bio}} = AREA_{crop}(m) * urb(m,t)$	GHS-POP[32] CROPGRIDS[24]; HYDE 3.3[23]
Atmospheric deposition	DEP	-	ISIMIP3a; Vishwakarma 2023[13,33]
Crop N removal	$CROP$	$\phi_{CROP}(j) = AREA_{harv}(m,j)$	CROPGRIDS[24]; HYDE 3.3[23]
N in grazing-derived animal products (demand term of forage removal, Eq. 29)	$N_{prod,graze}$	Same as manure deposited on pasture	Zhang et al. (2017); Global Pasture Watch[25,34]

Nitrogen mass balance and spatial scaling

The SCALES-Global N budget is constructed on a soil-surface, "as-applied" basis and is computed annually for every country from 1961 through 2023. We describe here our methods for estimating the components of the nitrogen mass balance and the construction of the three released accounting products: the cropland N surplus, the landscape N surplus, and the total anthropogenic N surplus. Unless otherwise noted, all budget components are first estimated at the national scale and then subsequently downscaled using component-specific spatial proxies.

Fertilizer

Country-scale synthetic N fertilizer application, $FERT_{appl}(i,t)$, is obtained from FAOSTAT and includes fertilizer applied to all agricultural land, including cropland and grazing land (defined here as managed pasture plus rangeland). Synthetic fertilizer is first partitioned between cropland and grazing land by multiplying by a country-specific cropland allocation coefficient, $c_f(i)$ derived from the FAO Cropland Nutrient Budget (**Table 2**):[35]

$$FERT_{crop,appl}(i,t) = FERT_{appl}(i,t) * c_f(i) \quad \text{(Eq. 3)}$$

$$FERT_{graze,appl}(i,t) = FERT_{appl}(i,t) * (1 - c_f(i)) \quad \text{(Eq. 4)}$$

For the 21 countries in which a material share of fertilizer is applied to grazing land, predominantly pasture-based livestock economies such as New Zealand (c_f = 0.10), Ireland (0.20), and the Netherlands (0.50), $c_f(i) < 1$; for all remaining countries $c_f(i) = 1$, such that all agricultural fertilizer is assigned to cropland. The c_f coefficients are held constant through time, as no global time-resolved cropland/grazing land fertilizer-allocation product is currently available. Although the cropland-grazing land partition has likely evolved over the 1961-2023 period, particularly in regions with rapidly changing livestock production systems, these coefficients preserve national fertilizer totals while determining the distribution of fertilizer inputs, and their associated ammonia volatilization losses, between cropland and grazing land.

Table 2. Country-specific cropland fraction coefficients, *cf(i),* used to partition national synthetic fertilizer consumption between cropland and grazing land. The cropland nitrogen budget uses *cf-N* values; *cf-P_2O_5* are also listed for reference. Values are taken from the FAO Cropland Nutrient Budget, Table 1.[35] For all countries not listed here, *cf = 1*, meaning that all fertilizer is assumed to be applied to cropland.

Country	ISO3	cf (N)	cf (P_2O_5)
Australia	AUS	0.90	0.70
Austria	AUT	0.90	0.90
Brazil	BRA	0.90	1.00
Canada	CAN	0.90	1.00
Chile	CHL	0.80	0.70
Finland	FIN	0.70	1.00
France	FRA	0.90	0.90
Germany	DEU	0.80	0.90
Ireland	IRL	0.20	0.30
Japan	JPN	0.80	1.00
Luxembourg	LUX	0.40	0.70
Morocco	MAR	0.90	1.00
Netherlands	NLD	0.50	0.90
New Zealand	NZL	0.10	0.10
Poland	POL	0.80	0.90
Slovenia	SVN	0.60	0.70
South Africa	ZAF	0.90	0.90
Switzerland	CHE	0.70	0.70
United Kingdom	GBR	0.80	0.70
United States	USA	0.80	1.00
Uruguay	URY	0.90	0.90

After the partitioning, fertilizer magnitudes are then converted from *applied* fertilizer to effective soil-surface inputs by subtracting field ammonia volatilization losses, as follows:

$$FERT_{crop,net}(i,t) = FERT_{crop,appl}(i,t) * (1 - fracGASF(i,t)) \quad \text{(Eq. 5)}$$

$$FERT_{mpast,net}(i,t) = FERT_{mpast,appl}(i,t) * (1 - fracGASF(i,t)) \quad \text{(Eq. 6)}$$

where *fracGASF(i,t)* is the fraction of applied N fertilizer that is lost to the atmosphere as a gas. The volatilization loss fractions differ by country and are calculated as follows:

$$fracGASF(i,t) = \frac{\sum_{p \in i} \; p_{use}(i,p,t) * N(p) * fracGASF(p)}{\sum_{p \in i} \; p_{use}(i,p,t) * N(p)} \quad \text{(Eq. 7)}$$

where $p_{use}(i,p,t)$ is the FAOSTAT reported use of product *p* in country *i* in year *t*, *N(p)* is the fractional N content of product *p*, and *fracGASF(p)* is the N volatilization rate of product *p*.

The 2019 IPCC Guidelines provide two synthetic-fertilizer loss fractions (0.15 for urea and urea-based products and 0.05 for all other synthetic nitrogen);[36] we apply the urea value directly and resolve the "other" category to individual products (**Table 3**) using the EMEP/EEA (2023) emission inventory guidebook[37] and the ammonia-volatilization synthesis of Bouwman et al.[38] This approach shows losses in nitrate-dominated markets (e.g., ~0.08 in the United States) and higher losses in urea-dominated ones (e.g., ~0.13 in India), relative to the global nitrogen-weighted default of ~0.11, which is applied where product-level data are unavailable.

Note that FAOSTAT tracks only agricultural fertilizer use; non-agricultural, domestic applications (lawns, gardens, turf, and other non-farm uses) are not reported and are consequently not accounted for here.

Table 3. Fertilizer-product NH_3-volatilization loss fractions (*FracGASF*) and nitrogen contents used to compute country-specific synthetic-fertilizer volatilization. Loss fractions are from the 2019 IPCC Guidelines,[36] the EMEP/EEA emission inventory guidebook,[37] and the ammonia-volatilization synthesis of Bouwman et al. .[38]

Fertilizer product	Item code	N content	FracGASF
Urea	4001	46%	0.15
Diammonium phosphate (DAP)	4022	18%	0.10
Other nitrogenous fertilizers, n.e.c.	4008	30%	0.10
Urea–ammonium nitrate solution (UAN)	4006	30%	0.08
Ammonium sulphate	4002	21%	0.08
Fertilizers, n.e.c.	4030	10%	0.08
NPK compound fertilizers	4021	15%	0.06
Other NK compounds	4026	13%	0.06
Other NP compounds	4024	18%	0.06
Ammonium nitrate (AN)	4003	34.5%	0.03
Calcium ammonium nitrate (CAN)	4004	27%	0.03
Monoammonium phosphate (MAP)	4023	11%	0.03
Anhydrous ammonia	4007	82%	0.02
Sodium nitrate	4005	16%	0.01
Potassium nitrate	4025	13.5%	0.01

Downscaling of Cropland Fertilizer: Developing the Φ_{FERT} term

Synthetic fertilizer applied to cropland is distributed using a crop-specific fertilizer proxy derived from the gridded fertilizer maps of Coello et al.,[12] which provide annual synthetic N fertilizer magnitudes at 5 arcmin resolution from 1961-2019 for 13 crop groups.

Coello et al.[12] construct these maps by applying a single, nationally predicted application rate for each crop group, country and year to their own gridded harvested-area map, which combines the EarthStat crop distribution for 2000[39] with annual HYDE 3.3 cropland area.[23] Within a country, the spatial pattern of the Coello fertilizer map therefore reflects differences in the harvested area of individual crops, but not differences in application intensity. Because the earlier EarthStat-based crop geography differs from

the updated CROPGRIDS crop-specific harvested-area maps (native 0.05°, circa 2020) used to allocate crop N removal in SCALES-global, applying the Coello spatial pattern directly would place fertilizer and harvest on two different crop maps and produce incorrect cell-level deficits. We therefore use Coello et al.[12] only to partition each country's cropland fertilizer among crop groups, and we allocate each group's share using the same CROPGRIDS harvested-area basis used for crop removal.

For crop group *j*, the national Coello fertilizer total in country *i* is obtained by summing the Coello fertilizer mass over the 5-arcmin cells *y* within the country:

$$F_c(j,i,t) = \sum_{y \in i} \quad d_{fert}(j,y,t) * AREA_{tot}(y) \qquad \text{(Eq. 8)}$$

where $d_{fert}(j,y,t)$ is the Coello areal mass density (kg N ha^{-1} of grid cell) and $AREA_{tot}(y)$ is the Coello 5-arcmin cell area. Cells that straddle a national border are apportioned between countries by their area overlap with the 0.05° grid. Each country's group total is then allocated to the 0.05° cells *m* within the country in proportion to that group's harvested area in year *t*, which is obtained, as for crop removal, by multiplying HYDE 3.3 cropland in year *t* by a static scaling factor:

$$FERT_{crop}(j,m,t) = F_c(j,i,t) * \frac{H(m,t)*s_j(m)}{\sum_{m\prime \in i} \quad H(m\prime,t)s(m\prime)}, s_j(m) = \frac{AREA_{harv}(j,m)}{H(m,2020)} \qquad \text{(Eq. 9)}$$

where $H(m,t)$ is HYDE 3.3 cropland in cell *m* and year *t*, and $s_j(m)$ is a scaling term calculated as the ratio of the group's CROPGRIDS total harvested area, $AREA_{harv}(j,m)$, to HYDE cropland in 2020, the CROPGRIDS reference year. The cropland area of each cell in year *t* is HYDE 3.3 cropland, $H(m,t)$; CROPGRIDS determines only how that cropland is divided among crops. Where HYDE reports no cropland in a cell in 2020, $s_j(m)$ is undefined and the cell is treated as having no harvested area, consistent with HYDE; the CROPGRIDS harvested area recorded in such cells is reallocated to the remaining cells of the country based on the national renormalization. In addition, where Coello reports fertilizer for group *j* but CROPGRIDS has no harvested area for that group anywhere in the country, the crop group fertilizer total is instead spread over the harvested area of all crops in the country. This fallback affects only 0.01% of global cropland fertilizer. After 2019, when the Coello dataset ends, the 2019 division of each country's fertilizer among crop groups is held fixed, while the national magnitudes continue to follow the FAOSTAT annual fertilizer data.

The proxy's per-cell fertilizer mass is then obtained by summing across crop groups:

$$\phi_{FERT_{crop}}(m,t) = \sum_j \quad FERT_{crop}(j,m,t) \qquad \text{(Eq. 10)}$$

The proxy is renormalized within each country so that the gridded total equals national cropland fertilizer (FAOSTAT agricultural use × cropland fraction). Coello et al. therefore determine only how a country's fertilizer is divided among crop groups, and the spatial distribution of fertilizer for individual crop groups follows CROPGRIDS. For the small number of island and territory countries that Coello does not cover (0.02% of global cropland), the NPKGRIDS crop-specific application-rate pattern is used instead.[40]

Downscaling of Grazing-Land Fertilizer

Synthetic fertilizer applied to pasture, $FERT_{mpast}(i,t)$ *is* distributed in proportion to the managed-pasture area within each grid cell, $AREA_{mpast}(m,t)$, as obtained from the Global Pasture Watch (GPW) managed-pasture base,[25] The 30-m resolution GPW managed pasture is aggregated to the SCALES-Global 0.05°

grid to obtain the managed-pasture area in each cell, $AREA_{mpast}(m,t)$, which is then used as the downscaling proxy:

$$\phi_{FERT_{mpast}}(m,t) = AREA_{mpast}(m,t) \quad \text{(Eq. 11)}$$

Livestock Manure N

Country-scale manure inputs are estimated separately for cropland and grazing land to reflect differences in livestock management and subsequent N losses. Manure N applied to cropland, $MAN_{crop,appl}(i,t)$, and grazing land, $MAN_{graze,appl}(i,t)$, are taken directly from FAOSTAT. FAOSTAT computes manure N production from national livestock inventories using IPCC Tier-1 excretion coefficients and management-system loss factors.[41,42] The Tier-1 coefficients specify default N-excretion rates (kg-N animal^{-1} yr^{-1}) by livestock species *j* and broad climatic region. For the cropland manure, FAOSTAT subtracts the N lost during housing and storage, i.e. ammonia and NO_x volatilization, leaching and runoff from housing and storage facilities, and N_2/N_2O emissions, using IPCC management-system loss fractions specific to each system and climate region.[41,42] The remaining N, after these management losses, is the manure applied to cropland soils.

Consistent with the soil-surface accounting framework adopted throughout SCALES-Global, field ammonia volatilization losses, which occur during and after application, are then removed from the cropland-applied manure:

$$MAN_{crop,net}(i,t) = MAN_{crop,appl}(i,t) * (1 - fracGASM_{crp}(i)) \quad \text{(Eq. 12)}$$

where $fracGASM_{crp}$ is the fraction of applied or deposited manure nitrogen that is lost to the atmosphere as a gas (0.21 for applied manure).[43]

For manure deposited directly onto pasture, range, and paddock, FAOSTAT assumes that excretion occurs directly on the field without housing or storage losses. Therefore, no management-stage losses are deducted from it, and the reported values represent manure N as deposited. To account for additional N volatilization occurring after manure falls on grazing land, we also subtract the field volatilization losses:

$$MAN_{graze,net}(i,t) = MAN_{graze,appl}(i,t) * (1 - fracGASM_{graze}(i)) \quad \text{(Eq. 13)}$$

where $fracGASM_{graze}$ is the fraction of deposited manure nitrogen that is lost to the atmosphere as a gas from pasture or rangeland ($fracGASM_{graze}$ = 0.10).[37,38]

Cropland Manure N Downscaling: Developing the $\Phi_{MAN_{crop}}$ term

The within-country spatial distribution of manure is derived from the 5-arcmin global gridded manure dataset of Zhang et al.,[31] which reconstructs annual manure production and cropland application for 1961–2014 using FAOSTAT livestock statistics, the Global Livestock Impact Mapping System (GLIMS), IPCC excretion coefficients, livestock production systems, and agro-ecological zones. Two Zhang layers are used: (1) the gridded manure applied to cropland is the spatial proxy for cropland manure, and (2) the difference between total manure production and cropland application is the proxy for grazing land manure. While Zhang et al. is used to set the within-country geography, all national manure magnitudes are taken from FAOSTAT. Because the Zhang reconstruction is itself based on FAOSTAT livestock statistics, its spatial patterns are consistent with the national activity data underlying the FAO Cropland Nutrient Budget while retaining the spatial detail of the gridded reconstruction.

Because the Zhang grid (5 arcmin) is coarser than the SCALES-Global grid (0.05°), **cropland manure** is redistributed within each Zhang cell onto the 0.05° grid in proportion to where cropland actually lies in year *t*, using the HYDE 3.3 cropland dataset.[23]

The manure mass in a Zhang cell *y* is obtained as

$$M_{man,crop}(y,t) = \delta_{man_{crop}}(y,t) * AREA_{tot}(y) \quad \text{(Eq. 14)}$$

where $\delta_{man_{crop}}$(y,t) is the manure-N application density (kg-N km^{-2} yr^{-1}) in the Zhang 5-arcmin cells and $AREA_{tot}(y)$ is the Zhang cell area. This mass is allocated to the 0.05° cells *m* within *y* in proportion to the HYDE cropland, $H(m,t)$:

$$MAN_{crop}(m,t) = M_{man,crop}(y,t) * \frac{H(m,t)}{\sum_{m\prime \in y} H(m\prime,t)} \quad \text{(Eq. 15)}$$

Where a Zhang cell contains no HYDE cropland in year *t*, the mass of manure N is instead distributed evenly by cell area. The resulting grid is then used as the proxy for distributing the country-scale crop manure data.

$$\phi_{MAN_{crop}}(m,t) = MAN_{crop}(m,t) \quad \text{(Eq. 16)}$$

Grazing-Land Manure N Downscaling: Developing the Φ_{graze} term

Grazing-land manure is downscaled like cropland manure, in that the Zhang 5 arcmin manure mass is redistributed onto the SCALES-Global 0.05° grid by area-conservative overlap, but it is downscaled to the broad GPW (managed pasture + rangeland) grazing base[25] rather than to cropland. The 30-m resolution GPW grazing land is aggregated to the SCALES-Global 0.05° grid to obtain the grazing-land area in each cell, $AREA_{graze}(m,t)$.

The mass of manure deposited to grazing land in each 5-arcmin Zhang residual cell is calculated as:

$$M_{man,graze}(y,t) = d_{MAN_{graze}}(y,t) * AREA_{tot}(y) \quad \text{(Eq. 17)}$$

It is then redistributed to the SCALES-Global 0.05° grid cells in proportion to $AREA_{graze}(m,t)$:

$$M_{man,graze}(m,t) = M_{man,graze}(y,t) * \frac{AREA_{graze}(m)}{\sum_{m\prime \in y} AREA_{graze}(m\prime)} \quad \text{(Eq. 18)}$$

Where a Zhang cell contains no grazing land, the mass of grazing-land manure N is instead distributed in proportion to non-cropland areas. The resulting grid is then used as the proxy for distributing the country-scale grazing-land manure data.

$$\phi_{MAN_{graze}}(m,t) = M_{man,graze}(m,t) \quad \text{(Eq. 19)}$$

The FAOSTAT national magnitudes are then distributed across grid cells in proportion to this proxy and normalized to conserve the national total.

Note that, as described in the "Spatial Downscaling Framework" section above, in a small number of countries, there is little or no mapped grazing land, meaning that grazing-land manure would be distributed to implausibly small areas of grazing land, or have no land at all available for distribution. In these cases, the grazing-land manure is reallocated to cropland.

Biological N Fixation

Country-scale biological N fixation, *BNF(i,t)* is estimated as the sum of symbiotic fixation by grain legumes, free-living fixation associated with rice and sugar cane, and pasture/forage-legume fixation on grazing land (managed pasture + rangeland).

$$BNF(i,t) = \sum_j \quad BNF_{crop,leg}(i,j,t) + \sum_j \quad BNF_{crop,free}(i,j,t) + BNF_{graze}(i,t) \qquad \text{(Eq. 20)}$$

Symbiotic fixation by grain legumes is estimated from harvested crop production:

$$BNF_{crop,leg}(i,j,t) = PROD(i,j,t) * c_N(j) * f_{bnf}(j) * wpf(j) \qquad \text{(Eq. 21)}$$

where $BNF_{crop,leg}(i,j,t)$ is the estimate of country-scale biological N fixation for country *i, for* crop *j,* in year *t*; *PROD(i,j,t)* is the FAOSTAT harvested production (tonnes) of harvested crop *j*, c_N(j) (kg-N tonne^{-1}) is the crop-specific N removal coefficient from the "World crop coefficients for UN FAO" dataset,[44] f_{bnf} (j) is the fraction of crop N considered to be derived from atmospheric fixation, and *wpf*(j) is the whole-plant factor relating grain N to the total plant N fixed.[45–47] The f_{bnf}(j) coefficient is region-specific for soybeans (0.78 for Brazil, 0.44 for Europe, 0.61 for the rest of the world) and group-based for pulses (**Table 4**).[48]

Table 4. Biological N fixation coefficients.

Crop (FAO item)	FAO code	%Ndfa (f_{bnf})	wpf	c_N (kg N t^{-1})
Soya beans	236	0.78 (Brazil) / 0.44 (Europe) / 0.61 (RoW)	1.90	55.9
Groundnuts, with shell	242	0.62	3.13	34.3
Beans, dry	176	0.38	4.42	42.2
Chick peas	191	0.62	4.42	27.3
Peas, dry	187	0.62	4.42	39.0
Lentils	201	0.62	4.42	35.8
Cow peas, dry	195	0.62	4.42	30.9
Bambara beans	203	0.62	4.42	25.7
Vetches	205	0.62	4.42	32.4
Pulses nes	211	0.62	4.42	26.1
Broad beans, horse beans, dry	181	0.74	4.42	30.2
Lupins	210	0.74	4.42	46.8
Pigeon peas	197	0.74	4.42	30.0
Rice (free-living)	27	— (area × 25 kg N ha^{-1})	—	—
Sugar cane (free-living)	156	— (area × 25 kg N ha^{-1})	—	—

Free-living and associative fixation associated with rice and sugar cane is estimated using the area-based coefficient adopted by the FAO Cropland Nutrient Budget:[35]

$$BNF_{crop,free}(i,j,t) = AREA_{crop,j}(i,j,t) * r_{free} \qquad r_{free} = 25\ kg\ N\ ha^{-1} \qquad \text{(Eq. 22)}$$

where r_{free} is the area-based coefficient value (25 kg-N ha^{-1} yr^{-1}) adopted by the FAO Cropland Nutrient Budget.

Pasture/forage-legume fixation on grazing land is estimated as an area-based input over managed pasture and rangeland, $AREA_{graze}(i,t)$:[49]

$$BNF_{graze}(i,t) = AREA_{graze}(i,t) * r_{graze} \quad \text{(Eq. 23)}$$

where r_{graze} (5 kg-N ha^{-1} yr^{-1}) is the area-averaged rate of symbiotic biological N fixation by forage and pasture legumes over grazing land. Because the legume fraction of grazing land is not spatially resolved, r_{graze} is applied as a coarse whole-pasture global average rather than a legume-specific rate: sown, legume-rich pastures fix on the order of 20–100 kg-N ha^{-1} yr^{-1} while extensive rangeland fixes very little, and the 5 kg-N ha^{-1} yr^{-1} value is chosen so that the global pasture/forage-legume fixation total is consistent with the estimates of Peoples et al.[50,51] This term contributes to the landscape and total surplus views only; it is excluded from the cropland (FAO-comparable) budget.

Note that fixation by forage legumes grown on arable land as hay or fodder, chiefly alfalfa and clover, is not quantified, because FAOSTAT does not report either their production or area. This omission likely understates cropland nitrogen inputs in regions with substantial legume hay cultivation (e.g., the US Upper Midwest, Argentina, and parts of Europe), a limitation shared with other FAOSTAT-based nutrient-budget products.[35]

BNF Downscaling: Developing the Φ_{BNF} terms

The legume and free-living fixation terms, $BNF_{crop,leg}(i,j,t)$ and $BNF_{crop,free}(i,j,t)$, and the pastureland term, $BNF_{graze}(i,t)$, are each distributed using the dasymetric approach of Eqs. 1-2. National cropland fixation is estimated separately for each nitrogen-fixing FAO crop item *j.* Each crop's national fixation is then distributed to that specific crop's harvested area in year *t*, the same proxy used for fertilizer and crop removal (Eqs. 9 and 28):

$$\phi_{BNF_{crop,leg}}(m,j,t) = \phi_{CROP}(m,j,t) \quad \text{(Eq. 24)}$$

$$\phi_{BNF_{crop,free}}(m,j,t) = \phi_{CROP}(m,j,t) \quad \text{(Eq. 25)}$$

$BNF_{graze}(i,t)$ is downscaled to the broad GPW (managed pasture + rangeland) grazing base, similar to grazing-land manure:

$$\phi_{BNF_{graze}} = AREA_{graze}(m,t) \quad \text{(Eq. 26)}$$

Crop and Forage N Removal

Crop N removal

Nitrogen removal by crop *j*, $CROP(i,j,t)$ (kg-N yr^{-1}), represents the nitrogen contained in harvested crops in country i and year t. It is calculated as follows:

$$CROP(i,j,t) = (PROD(i,j,t) * c_N(j)) \quad \text{(Eq. 27)}$$

where $PROD(i,j,t)$ (t yr^{-1}) is FAOSTAT production of crop *j* in country *i* in year *t* in (tonnes yr^{-1}), and c_N(j) (kg-N tonne^{-1}) is the crop-specific N removal coefficient from the "World crop coefficients for UN FAO"

dataset, with the following exceptions.[44] Specifically, eight world values in the crop coefficient dataset were corrected before use because each averaged in a source that does not correspond to the fresh harvested product reported by FAOSTAT: sugar cane (3.674 to 0.984 kg N t^{-1}; a bagasse entry excluded), papayas (79.55 to 0.64; an implausible 119 kg N t^{-1} value excluded), coconuts (15.37 to 6.24; a non-fresh-nut value and a misfiled silk-cocoon entry excluded), oil palm fruit (3.916 to 2.90; palm kernel and oil values excluded, leaving fresh fruit bunches), mushrooms (9.28 to 3.20; dried mushrooms excluded), mangoes (4.04 to 1.07; an outlier excluded), plantains (3.98 to 1.29; an outlier excluded), and cassava (2.47 to 2.20; dried cassava excluded). After removal values are calculated, the results are summed across ~150 crop commodities to estimate the CROP nitrogen removal at the country scale.

Crop Downscaling: Developing the Φ_{CROP} term

Crop N removal is downscaled crop-by-crop rather than from a single national value. The national removal of each crop *j*, *CROP(i,j,t)*, is allocated to grid cells in proportion to that crop's harvested-area proxy, and the per-crop grids are summed. The spatial proxy for crop j is its gridded harvested area in year *t*, obtained in the same way as for the fertilizer crop groups (Eq. 9):

$$\phi_{CROP}\ (m, j, t) = s_j(m) * H(m, t) \qquad \text{(Eq. 28)}$$

where $s_j(m)$ is the ratio of the CROPGRIDS harvested area of crop *j* to HYDE cropland in 2020. Each crop's national removal is thus conserved within the country, and grid-cell totals are summed across all crop types.

Forage N Removal

Grazing land forage nitrogen removal, *FORAGE(i,t)*, is typically a weakly constrained output term in landscape-scale nitrogen budgets. Existing approaches generally estimate grazing land removal either from supply, using grazing land area or productivity multiplied by an assumed nitrogen uptake rate,[52,53] or from demand, using livestock numbers, feed requirements, and animal production. Regional grazing land budgets often combine these approaches,[54–56] whereas many global crop nitrogen budgets omit grazing land entirely.[9,34,35]

SCALES-global adopts a hybrid approach, following the logic of the US county-scale TREND-nitrogen and gridded gTREND datasets[10,17] With this approach, grazing-livestock forage demand is bounded by a pasture supply ceiling, and the forage N removed from grazing land is taken as the lesser of the nitrogen taken up by grazing animals and the maximum nitrogen the pasture can yield. The constraint is evaluated at the grid cell *m*, the scale at which grazing land area and livestock density are resolved:

$$FORAGE(m,t) = min[MAN_{graze}(m,t) + N_{prod,graze}(m,t),$$
$$p_{managed} * AREA_{mpast}(m,t) + p_{range} * (AREA_{graze}(m,t) - AREA_{mpast}(m,t))] \qquad \text{(Eq. 29)}$$

The national value is the sum over the cells of each country. Here, the demand term ($MAN_{graze}(m,t) + N_{prod,graze}(m,t)$) represents the nitrogen consumed by grazing animals and either returned to pasture as dung and urine or exported in animal products. $MAN_{graze}(m,t)$ is the gridded pasture-deposited manure of Eq. 19. $N_{prod,graze}$ represents nitrogen exported in grazing-derived animal products, estimated from FAOSTAT production of ruminant meat, milk, and wool multiplied by product-specific nitrogen contents and distributed within each country on the grazing proxy of Eq. 30. The supply ceiling is the grazing land area of the cell multiplied by an area-based pasture N-uptake rate that depends on grazing land type. Managed-pasture area, $AREA_{mpast}$, and the grazing base, $AREA_{graze}$, are the Global Pasture Watch areas described above,[25] so that their difference is natural and semi-natural rangeland, with $p_{managed}$ = 200 and

p_{range} = 30 kg N ha^{-1} yr^{-1}, the midpoints of measured N yields of managed temperate grassland (approximately 100–300 kg N ha^{-1} yr^{-1}) and extensive rangeland (10–50 kg N ha^{-1} yr^{-1}).[57,58]

This supply-ceiling approach caps removal in intensively stocked or feed-supplemented systems, where grazing demand would otherwise exceed what the pasture can produce. Following the TREND approach,[10,17] bounding demand by supply prevents over-attributing forage removal to grazing land in confined-feeding systems. Note that the forage removal term enters the landscape and total surpluses only; the cropland surplus is unchanged.

Forage Downscaling: Developing the Φ_{PROD} term

Because forage removal is evaluated per cell (Eq. 29), only nitrogen in grazing-derived animal products, $N_{prod,graze}(i,t)$, is downscaled from a national value. Its proxy, $\phi_{PROD}(m,t)$, is the proxy for manure deposited to grazing land (Eq. 19), which uses the Zhang et al. grazing land manure-N deposition rate multiplied by the grazing-land area.

$$\phi_{PROD}(m,t) = \phi_{MAN_{graze}}(m,t) = M_{man,graze}(m,t) \quad \text{(Eq. 30)}$$

Atmospheric deposition

As mentioned above, atmospheric deposition is the one term not downscaled by this normalized-proxy method. National atmospheric N deposition, *DEP,* is derived from two products. The national cropland deposition rate estimates of Vishwakarma et al.[13] set the deposition levels for cropland (1961-2022), and the gridded ISIMIP3a atmospheric deposition field provides the rates on non-cropland land (1961-2021).[33] Both datasets are held at the last available year to provide SCALES-Global coverage to 2023. We use Vishwakarma's WL product[13] to set the cropland deposition levels because, in their station-network evaluation, the Wang-based fields most closely reproduced observed deposition in the high-deposition regions of China and East and South Asia, where global deposition products diverge most strongly and where cropland deposition is largest.

The deposited N mass in each grid cell *m* is the sum of cropland and non-cropland deposition:

$$DEP\ \ (i,m,t) = DEP_{crop}(i,m,t) + DEP_{non-crop}(i,m,t) \quad \text{(Eq. 31)}$$

$$DEP_{crop}(i,m,t) = W(i,t) * A_{crop}(i,m,t) \quad \text{(Eq. 32)}$$

$$DEP_{non-crop}(i,m,t) = I(m,t) * (A_{tot}(i,m,t) - A_{crop}(i,m,t)) \quad \text{(Eq. 33)}$$

where $W(i,t)$ is the cropland deposition rate from the WL product of Vishwakarma et al.,[13] $I(m,t)$ is the ISIMIP3a deposition rate, $A_{crop}(i,m,t)$ is the cropland area (HYDE 3.3 cropland, $H(m,t)$), and $A_{tot}(i,m,t)$ is the total area in grid cell *m*. $W(i,t)$ is a single national rate applied uniformly to all cropland area in country *i,* whereas $I(m,t)$ varies cell-by-cell and is taken from the ISIMIP3a deposition grid (0.5°, total N, regridded to the SCALES-Global 0.05° grid).

The country-scale cropland deposition magnitude is the sum of $DEP_{crop}(i,m,t)$ for grid cells for country *i;* the landscape and total surplus budgets use the sum of $DEP(i,m,t)$ over those same cells.

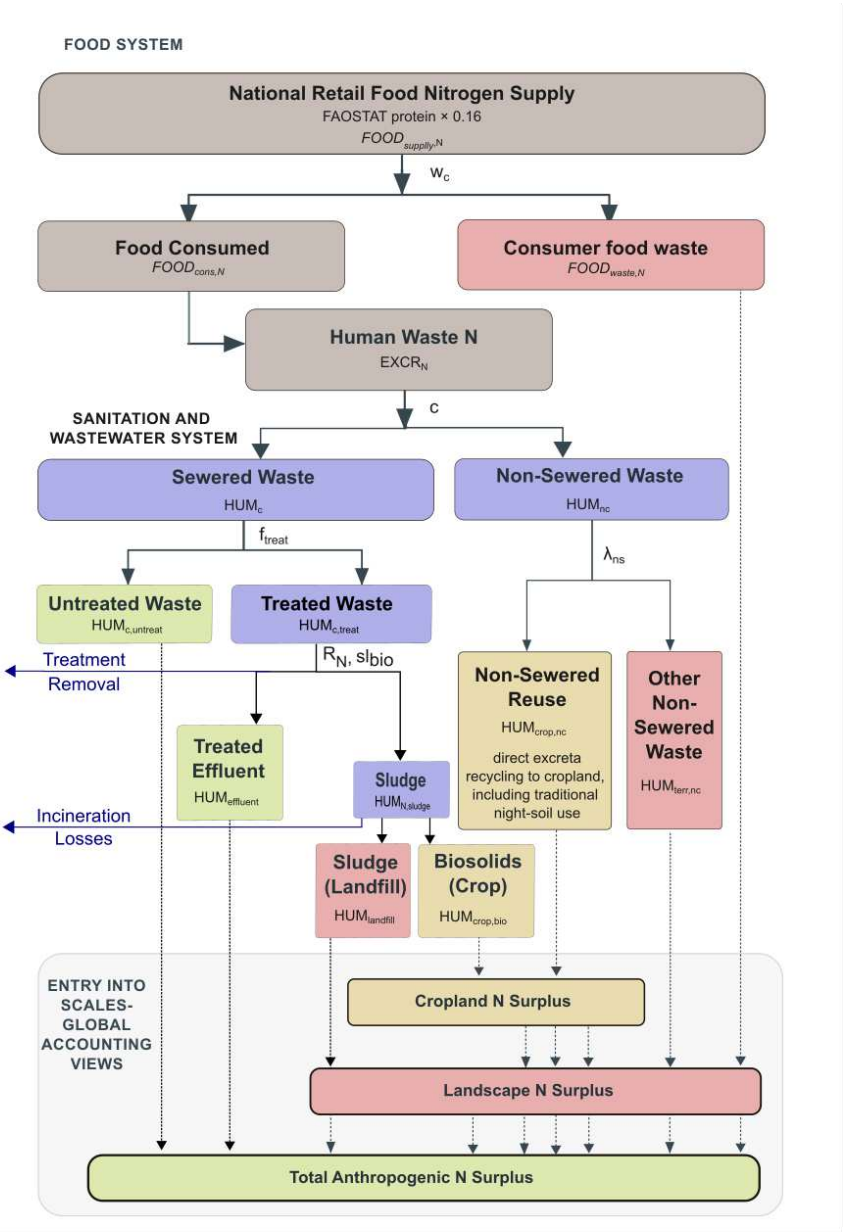
Note that SCALES-Global treats both oxidized (NOy) and reduced (NHx) atmospheric deposition as external inputs to the cropland surface. This differs from net-anthropogenic-nitrogen-input (NANI)-style

accounting,[59] in which only oxidized deposition is treated as a new input, while reduced N (NH_3/NH_4^+) is assumed to be internally cycled, on the reasoning that ammonia volatilized from fertilizer and manure within a watershed is redeposited locally.[60] Although this assumption may be appropriate for watershed-scale budgets, it is less suitable for national accounting because substantial fractions of atmospheric NHx may be transported across regional and national boundaries before deposition.[61,62] Consequently, much of the reduced N deposited onto a country's cropland may originate outside that country's agricultural system and therefore may represent a genuine external input. Including total deposition in the budget is also consistent with the soil-surface accounting framework used by the Eurostat/OECD Gross Nitrogen Balance[63] and with the FAO Cropland Nutrient Budget[35] against which we benchmark.

Human N Fluxes

The human food system redistributes reactive nitrogen through food consumption, waste generation, wastewater treatment, and environmental release (**Figure 3**). SCALES-Global reconstructs these pathways from a common mass balance based on the national food supply, population, sanitation access, and wastewater-treatment data. Retail food nitrogen is first partitioned between food consumed and consumer food waste. Nitrogen consumed by humans is assumed to be excreted and is then partitioned between sewered and non-sewered sanitation systems. Within sewered systems, nitrogen enters one of three pathways: (1) treatment removal, (2) effluent discharged to surface water, and (3) sludge. Nitrogen in sludge is either recycled to cropland as biosolids, lost to the atmosphere through incineration, or disposed of in landfills or applied to other land. Within non-sewered systems, nitrogen is divided between direct reuse on cropland and other releases to the terrestrial environment.

These human N inputs are included differently across the three SCALES-Global surplus estimates. The cropland surplus includes human-derived nitrogen intentionally returned to cropland through biosolids and non-sewered reuse. The landscape surplus includes the cropland surplus and additionally accounts for human-derived nitrogen disposed of or applied elsewhere on land, including sludge and food waste disposed of in landfills or applied to non-cropland soils. The total anthropogenic surplus includes the landscape surplus and further accounts for nitrogen discharged in treated wastewater effluent.



559

560 **Figure 3. Human nitrogen pathways represented in SCALES-Global.** Conceptual diagram showing how
561 human-derived nitrogen (N) is partitioned from the national food supply through consumption, food
562 waste, excretion, sanitation, wastewater treatment, biosolids management, and environmental return
563 pathways. Recycled human N applied to cropland contributes to the cropland, landscape, and total
564 anthropogenic surplus views; human N returned to the broader terrestrial landscape contributes to the
565 landscape and total anthropogenic views; and wastewater-derived N discharged as effluent to surface
566 water contributes only to the total anthropogenic view. Treatment removal represents N removed
567 during wastewater treatment processes before environmental discharge.

National Retail Food Nitrogen Supply

The national per-capita protein supply, $PROTEIN(i,t)$, from FAOSTAT[49] is converted to a retail food nitrogen supply, $FOOD_{supply,N}$(i,t) (kg-N yr^{-1}), assuming that protein contains 16% nitrogen by mass.[64]

$$FOOD_{supply,N}(i,t) = PROTEIN(i,t) * 0.16 * 365/1000 * POP(i,t) \quad \text{(Eq. 34)}$$

Because the FAOSTAT protein supply is reported at the retail stage, it includes food that is subsequently discarded by retailers and consumers. We partitioned the national retail nitrogen supply into consumed food and consumer food waste, $FOOD_{waste,N}(i,t)$ (kg-N yr^{-1}), using the regional retail- and household-level food-loss fractions, $w_c(i,t)$, of Morée et al.[20]

$$FOOD_{waste,N}(i,t) = FOOD_{supply,N}(i,t) * w_c(i,t) \quad \text{(Eq. 35)}$$

These fractions are specified as a uniform 6.7% in 1900 and then increase region-specifically from 10-26% by 2010. (Table 5); for intermediate years, we linearly interpolate between those endpoints, so mid-century food waste is lower than today's, and food waste increases over time. Nitrogen in food waste is considered to enter the landscape through the solid-waste stream (landfill, compost, disposal).

Table 5. Regional post-retail (retail + household) food-loss fractions, w_c. Values from Morée et al.[20] The food-loss fraction, w_c, is the sum of retail and household losses. In 1900, values are assumed to be uniform across regions. By 2010, they are region-specific. Intermediate years are linearly interpolated between the 1900 and 2010 endpoints and held constant outside that range.

Region	w_c, 1900 (%)	Retail loss, 2010 (%)	Household loss, 2010 (%)	w_c, 2010 (%)
Europe (including Russia)	6.7	4.8	13.4	18.2
North America and Oceania	6.7	5.1	21.1	26.2
Industrialized Asia	6.7	5.4	10.0	15.4
Sub-Saharan Africa	6.7	8.3	1.9	10.2
North Africa, West and Central Asia	6.7	6.9	6.6	13.5
South and Southeast Asia	6.7	8.1	3.0	11.1
Latin America	6.7	6.3	5.7	12.0

The consumed portion of the retail food nitrogen supply is assumed to enter the N waste stream as human excreta, calculated as the complement of Eq. 35:

$$FOOD_{cons,N}(i,t) = FOOD_{supply,N}(i,t) * (1 - w_c(i,t)) \quad \text{(Eq. 36)}$$

Assuming that humans exist in an approximate N balance, such that N excreted closely matches N ingested, the food N consumed is assumed to equal total human N excretion, $EXCR_N(i,t)$ (kg-N yr^{-1}) as:

$$EXCR_N(i,t) = FOOD_{cons,N}(i,t) \quad \text{(Eq. 37)}$$

Sewered and Non-Sewered Waste Pathways

Excreted nitrogen, $EXCR_N(i,t)$, is partitioned into sewered (collected) and non-sewered (non-collected) pathways:

$$HUM_c(i,t) = EXCR_N(i,t) * c(i,t) \quad \text{(Eq. 38)}$$

$$HUM_{nc}(i,t) = EXCR_N(i,t) * (1 - c(i,t)) \quad \text{(Eq. 39)}$$

where $c(i,t)$ is the time-varying collection fraction, i.e. the share of a country's human waste that is centrally collected or sewered. We assemble $c(i,t)$ for every country from a four-tier priority cascade: (1) observed national statistics; (2) country-specific anchors; (3) a region-calibrated GDP model for countries with no reported observations; and (4) a regional median for the few remaining small states. Tiers 1-3 share a common logistic development weight, $s(i,t)$, which lets sewerage scale with income where observations are sparse. We define the development weight here, and then describe each tier in more detail.

Development Weight. Following the space-for-time approach of Van Drecht et al.[19] and Van Puijenbroek et al.,[65] the development weight, $s(i,t)$, is a logistic function of per-capita GDP, $G(i,t)$:[66]

$$s(i,t) = \frac{1}{(1+exp[-K(log_{10}\, G(i,t) - X_0)])} \quad \text{(Eq. 40)}$$

where $G(i,t)$ is per-capita GDP in Maddison's units (2011 international $). The logistic parameters X_0 and K are obtained by fitting the sewer-connection fraction to per-capita GDP across countries for the years 1970-2021 (~1,540 country-years for 100 countries, r = 0.75), giving a midpoint GDP per capita value of ~$8,800 ($X_0$ = 3.95 and steepness K = 2.62). This single weight drives all three GDP-based operations below: the back-cast and forward-cast of the observed/anchor series (tiers 1-2) and the region-calibrated model (tier 3).

Country Tiers for Development of Collected/Sewered Fraction:

1. **Observed national statistics (120 countries)**. Where reported, $c(i,t)$ is taken from the United Nations Statistics Division series "population connected to a wastewater collecting system" (share of the resident population; 119 countries, 2000-2021; the United States is covered only by the OECD series),[67] extended by the OECD "total public sewerage" series (share of the population; 51 countries, 1970-2024),[68] which supplies observed values back to 1970 for 31 mostly-OECD country histories and agrees with the UNSD series to $r \approx 0.99$ over the overlapping years. The UNSD value takes precedence where the two overlap; the OECD series fills earlier years. Observations are used as reported, and interior gaps between reported years are filled by linear interpolation. For the extrapolated tails prior to and after the observed and reported data, values are scaled symmetrically from the nearest observation by the development weight ratio.

$$c(i,t) = c(i,t_0) * \frac{s(i,t)}{s(i,t_0)} \qquad t < t_0 \qquad \text{(Eq. 41)}$$

$$c(i,t) = c(i,t_1) * \frac{s(i,t)}{s(i,t_1)} \qquad t > t_1 \qquad \text{(Eq. 42)}$$

where t_0 and t_1 are the first and last reported years.

2. **Country-specific anchors (≈50 additional countries).** For countries absent from the UNSD series, we compiled the sewer-connection fraction from country- or region-specific sources, including national water and sanitation strategic plans, Inter-American Development Bank utility surveys, WHO Eastern Mediterranean regional wastewater overviews, the AMCOW / African Water Association sanitation white paper, and peer- reviewed national studies (**Supplemental Table S1**). Each anchor is treated as a single observation and scaled by the development weight in both directions from its reference (Eq.s 41-42), exactly as for the tier 1 countries. Five low-income countries that lack both a data source and a per-capita GDP record (South Sudan, Eritrea, Papua New Guinea, Timor-Leste, Bhutan) are assigned a documented regional- analogy value (all in the 0.005–0.05 range; **Supplemental Table S1**).

3. **Region-calibrated GDP model.** For countries with neither an observation nor an anchor but with a per-capita GDP record, *c(i,t)* is estimated from a global sewerage-income logistic scaled to the country's world region:

$$c(i,t) = min[\, s(i,t) * m_r \,, 1\,] \qquad \text{(Eq. 43)}$$

where m_r is a region-specific calibration factor equal to the median ratio of observed *c(i,t)* to *s(i,t)* across each region's observed members (**Table 6**). m_r is less than 1 where a region has less sewerage than income alone would predict (Sub-Saharan Africa, m_r = 0.17) and greater than 1 where it has more (Middle East, 1.15; Western Europe, 1.10), so the model reproduces regional infrastructure patterns that income cannot capture on its own.

4. **Region median.** The remaining small-island states and territories that have neither a reported, observed value nor a per-capita GDP record are assigned the median observed *c(i,t)* value for their world region.

Because centralized sewerage is durable infrastructure that is not typically decommissioned when incomes fall, *c* is constrained to be non-decreasing through time. Accordingly, for extrapolated years, a decline in per-capita GDP holds connection at its previous level rather than lowering it. Reported observations are retained as published, so any genuine decline in the measured record is preserved (**Figure 4**).

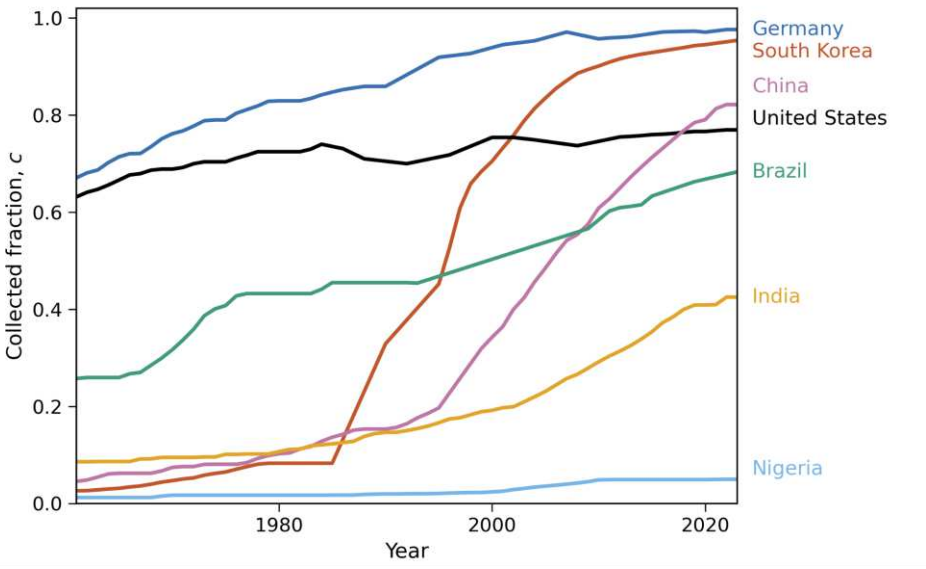


Figure 4. Reconstructed collected (sewer-connection) fraction c(i,t), 1961–2023, for representative countries. Countries span early/high (Germany, US), rapid developers (South Korea, China), intermediate (Brazil), and low-connection (India, Nigeria) trajectories.

Sewered Wastewater and Biosolids

Sewered waste can undergo primary, secondary or tertiary treatment, $HUM_{c,treat}(i,t)$, or can be discharged to the environment untreated, $HUM_{c,untreat}(i,t)$:

$$HUM_{c,treat}(i,t) = HUM_c(i,t) * f_{treat}(i,t) \quad \text{(Eq. 44)}$$

$$HUM_{c,untreat}(i,t) = HUM_c(i,t) * (1 - f_{treat}(i,t)) \quad \text{(Eq. 45)}$$

where $f_{treat}(i,t)$, is the fraction of the collected wastewater that receives treatment. We build $f_{treat}(i,t)$ from observed national statistics:

$$f_{treat}(i,t) = \frac{\tau(i,t)}{c(i,t)} \quad \text{(Eq. 46)}$$

where $\tau(i,t)$ is the fraction of the population connected to centralized wastewater treatment and *c(i,t)* is the fraction simply connected to a centralized collection system.[67,69] Both terms refer to the centralized/public system, and treatment here excludes independent, on-site facilities such as septic systems. On-site sanitation is instead represented in the non-sewered pathway (1-*c(i,t)*). $\tau(i,t)$ is reconstructed using the same scaling trajectory as that used for *c(i,t)* (Eqs. 40–42).

For countries without an observed treatment series, $f_{treat}(i,t)$ is set to a region-specific treated-to-collected ratio, τ_r , calculated as the median of τ/c across each region's observed members (countries reporting both collection and treatment) (**Table 6**). For high-income countries absent from both reporting programmes (for example Taiwan and Macao, which the UN and OECD series do not cover), τ_r is floored at the mean treated-to-collected ratio observed among high-income countries (0.94; n = 29

countries with GDP ≥ $25,000, median 1.0), reflecting that nearly all collected wastewater is treated at high incomes.

Table 6. Regional calibration factors for the human-waste collection and treatment model.

IPCC-Nex region	n (obs. c)	Median obs. c	m_r	n (obs. treat.)	τ_r (= f_{treat})
Africa	46	0.040	0.17	12	1.00
Asia	20	0.425	0.71	8	0.73
Eastern Europe	19	0.722	1.00	8	0.95
Indian Subcontinent	5	0.350	0.92	2	0.32
Latin America	30	0.422	0.97	17	0.81
Middle East	15	0.810	1.15	10	0.99
North America	3	0.755	0.93	3	0.99
Oceania	6	0.235	1.05	2	0.21
Western Europe	23	0.951	1.09	19	1.00

The untreated waste is discharged to receiving waters as a component of the total N surplus. The nitrogen that does enter treatment, $HUM_{c,treat}(i,t)$ is removed from the wastewater with a nitrogen removal efficiency of $R_N(i,t)$. We set this effective removal efficiency by ramping between the primary- and tertiary-treatment anchors of Van Drecht et al. (2009)[19] and Morée et al. (2013),[20] 0.10 and 0.80, respectively, with the country-level development weight, $s(i,t)$ (Eq. 40):

$$R_N(i,t) = R_{low} + (R_{high} - R_{low}) * s(i,t) \quad \text{(Eq. 47)}$$

with R_{low} = 0.10 and R_{high} = 0.80.

Treatment N Removal

The *treated* nitrogen is then split three ways: one fraction, $HUM_{den}(i,t)$, is denitrified and released to the atmosphere; another fraction is discharged to receiving waters as part of the treated effluent, $HUM_{effluent}(i,t)$; and the remaining solid fraction becomes sewage sludge, $HUM_{sludge}(i,t)$, :

$$HUM_{den}(i,t) = max\,[0, R_N(i,t) - sl_{bio}] * HUM_{c,treat}(i,t) \qquad sl_{bio} = 0.18 \quad \text{(Eq. 48)}$$

$$HUM_{effluent}(i,t) = (1 - R_N(i,t)) * HUM_{c,treat}(i,t) \quad \text{(Eq. 49)}$$

$$HUM_{sludge}(i,t) = min\,[sl_{bio}, R_N(i,t)] * HUM_{c,treat}(i,t) \qquad sl_{bio} = 0.18 \quad \text{(Eq. 50)}$$

where sl_{bio} is the fraction of the incoming treated nitrogen that remains with the sludge/solids portion of the waste–a fraction that has been shown to remain roughly constant (0.18) from basic to advanced treatment.[70–72] At the lowest treatment intensities, $R_N(i,t)$ is less than sl_{bio} (primary settling removes solids but results in no biological removal); here, the min/max terms set $HUM_{den}(i,t)$ to zero and put all removed nitrogen in the sludge, which is the physically correct limit for settling-only plants. This structure reproduces observed patterns, producing a roughly stable sludge fraction alongside growing

gaseous removal fractions as treatment intensifies, as seen in the closed full-scale nitrogen balances of Sagberg et al. (44% gas, 26% effluent, 16% biosolids at the tertiary VEAS plant),[73], Dai (23%, 51%, 26% in a secondary removal plant),[72] and Shi (48%, 40%, 12% activated sludge with denitrification).[74]

Recovered Sludge

The recovered sludge nitrogen, $HUM_{sludge}(i,t)$, is disposed of via three pathways according to country-specific sludge-management practices: (1) land application to cropland as biosolids, (2) landfill, and (3) incineration. Partitioning occurs in SCALES-Global with national-scale routing fractions: $a_{ag}(i,t)$, $a_{lf}(i,t)$, $a_{inc}(i,t)$, which sum to one.

$$HUM_{crop,bio}(i,t) = a_{ag}(i,t) * HUM_{sludge}(i,t) \quad \text{(Eq. 51)}$$

$$HUM_{landfill}(i,t) = a_{lf}(i,t) * HUM_{sludge}(i,t) \quad \text{(Eq. 52)}$$

$$HUM_{inc} \quad (i,t) = a_{inc}(i,t) * HUM_{sludge}(i,t) \quad \text{(Eq. 53)}$$

Disposal shares were compiled individually for 121 countries (every country with a non-negligible sewered, treated population), as anchor-year trajectories over 1961-2023 from national and regional biosolids statistics (Supplemental Table S2). These country-specific compilations vary in evidentiary strength, and each is assigned a confidence flag in the supplemental table. Fractions are linearly interpolated between anchor years and held flat outside the anchor range. The trajectories capture major documented regime shifts in wastewater management, e.g. Germany's increase in incineration from 5% in the 1960s to 81% in 2023 under the 2017 Sewage Sludge Ordinance,[75,76] the Dutch phase-out of agricultural application after the mid-1990s, the end of sea disposal of waste in 1998 in the United Kingdom,[75,76] and the shift in the United States away from ocean dumping after 1988.[77] Countries without country-specific data are assigned a global-average default mix, and the routing fractions are held constant through time. Incinerated sludge volatilizes essentially all of its nitrogen to the gas phase during combustion (N_2, NO_x, N_2O).[78,79]

Cropland biosolids, $HUM_{crop,bio}(i,t)$, contribute to the cropland, landscape, and total anthropogenic surpluses. Landfilled sludge, $HUM_{landfill}(i,t)$, is retained in the terrestrial landscape and therefore enters the landscape and total surpluses. Treatment denitrification, $HUM_{den}(i,t)$, and the off-gassing from incineration, $HUM_{inc}(i,t)$, leave the terrestrial system and are excluded from all surplus magnitudes.

Non-Collected Human Waste

A fraction, $\lambda_{ns}(i,t)$, of the non-sewered human N waste is recycled directly to cropland as "night soil," a practice with deep historical roots in East and Southeast Asia and, more sparsely, elsewhere.

$$HUM_{crop,nc}(i,t) = \lambda_{ns}(i,t) * HUM_{nc}(i,t) \quad \text{(Eq. 54)}$$

The $\lambda_{ns}(i,t)$ values are reconstructed by country using a two-tier approach. Tier 1 includes countries with well-documented national trajectories of human-excreta recycling to cropland (**Table 7**).[80,81] For these countries, $\lambda_{ns}(i,t)$ is specified from published trajectories of excreta recycling and linearly interpolated between anchor years (see Table 7).[82–84] Because the trajectories are typically reported as fractions of

total excreted N (R_{lit}), we express them in relation to the entire non-sewered pool and bound them at one,

$$\lambda_{ns,T1}(i,t) = min\left[\frac{R_{lit}(i,t)}{(1-c(i,t))}, 1\right] \quad \text{(Eq. 55)}$$

so that night soil never exceeds total non-sewered excreta N. For the eras when these countries recycled heavily, c ≈ 0 and λ_ns ≈ $R_{lit}(i,t)$. These trajectories implicitly reflect each country's sanitation transition, urbanization, fertilizer availability, and changing agricultural reuse practices.[85,86]

Tier 2 includes all remaining countries. For these countries, $\lambda_{ns}(i,t)$ is calculated as follows:

$$\lambda_{ns,T2}(i,t) = C_{region}(i) * M_{1950} \qquad M_{1950} = 0.15 \quad \text{(Eq. 56)}$$

reconstructed from the regional, early-20th century recycling classes of Morée et al.,[20] scaled to modern intensity. Regional recycling classes are specified as none, low, medium, and high reuse, corresponding to 0, 10, 40, and 70% of excreted nitrogen in 1900. Following Morée et al.,[20] the non-sewered recycling fraction is reduced to 15% of its 1900 value by 1950 and held constant thereafter.

The remainder of the non-sewered waste, human excreta that is neither collected for treatment nor deliberately recycled to cropland, is deposited to the wider landscape through pit latrines, septic systems, and open defecation. It is the complementary share of $HUM_{nc}(i,t)$:

$$HUM_{terr,nc}(i,t) = (1 - \lambda_{ns}(i,t)) * HUM_{nc}(i,t) \quad \text{(Eq. 57)}$$

$HUM_{terr,nc}(i,t)$ enters the landscape and total nitrogen surplus calculations, but not the cropland surplus, as the nitrogen is dispersed across the settled landscape rather than applied as a managed agronomic input.

Table 7. Tier-1 documented night-soil (human-excreta) recycling trajectories used in SCALES-Global.

Country (ISO3)	Basis / key sources	Trajectory and drivers	Representative $\lambda_{ns,T1}$ (anchor year(s) : value)
China (CHN)	*Gu et al. 2015; Starck et al. 2024*[16,87]	Near-universal pre-1980 (rural ~0.95, urban ~0.90). Declines with chemical fertilizer, sewerage, and urbanization.	1965–1980: 0.94; 1990: 0.82; 2000: 0.64; 2010: 0.28; 2020: 0.12
Japan (JPN)	*King 1911; Shirai et al. 2023; Noda 2024*[80,82,83,88]	Near-universal shimogoe recycling; steep 1960s collapse (chemical fertilizer, parasite-control campaigns, sewerage).	1961: 0.90; 1965: 0.72; 1980: 0.15; 1990: 0.05; 2000: 0.03; 2010–2020: 0.02
South Korea (KOR)	*King 1911; Starck et al. 2024*[82,87]	Earlier and steeper than China (1969 anti-parasite campaign; fastest-on-record urbanization and sewerage).	1961: 0.90; 1980: 0.55; 1990: 0.30; 2000: 0.10; 2010: 0.04; 2020: 0.02
North Korea (PRK)	*FAO/WFP CFSAM 2012; Radio Free Asia 2019–2024*[89–92]	U-shaped / non-declining: the post-1990 collapse of chemical-fertilizer supply drove a state-mandated return to human-excreta recycling (the "manure battle" quotas).	1961: 0.85; 1980: 0.45 (trough); 1990: 0.50; 2000: 0.75; 2010–2020: 0.80
Vietnam (VNM)	*Jensen et al. 2008; King 1911*[82,93–95]	Strong north/central reuse (Red River Delta double-vault composting; Mekong fish-pond latrines), weighted	1965: 0.70; 1980: 0.68; 1990: 0.58; 2000: 0.42; 2010: 0.30; 2020: 0.18

Country (ISO3)	Basis / key sources	Trajectory and drivers	Representative $\lambda_{ns,T1}$ (anchor year(s) : value)
		down by the aquaculture-oriented south.	
Taiwan (TWN)	*King 1911; Japanese-colonial organized-collection records*[82]	Chinese night-soil belt plus Japanese-era organized collection; steep industrial-era collapse. Estimate calibrated to JPN/KOR.	1961: 0.80; 1975: 0.55; 1985: 0.20; 2000: 0.05; 2020: 0.02
Hong Kong (HKG)	*Hong Kong EPD; New Territories aquaculture / night-soil-trade histories*[96]	Urban night-soil trade to New Territories farms and the Pearl River Delta, plus excreta-fed fishponds; collapse by the 1980s. Estimate.	1961: 0.65; 1975: 0.30; 1990: 0.05; 2020: 0.01
Nepal (NPL)	*WaterAid (ecosan Nepal); Himalayan dry-toilet composting literature*[97]	Persistent hill / Kathmandu-Valley composting-toilet tradition; slow decline with some ecosan revival. Estimate.	1961: 0.35; 1980: 0.30; 2000: 0.22; 2020: 0.15

Downscaling of Human Waste: Developing the Φ terms

The national human-derived nitrogen terms are allocated with the same normalized-proxy method as every other component (Eqs. 1–2; Table 1). $FOOD_{waste,N}(i,t)$ is treated here as a landscape input, analogous to the $FOOD_{waste}$ terms of Van Meter et al. and Byrnes et al. for phosphorus,[98,99] but kept here as a single land-applied input rather than routed to a separate landfill compartment. Consumer food waste is distributed according to population:

$$\phi_{FOOD_{waste}} = POP(m,t) \quad \text{(Eq. 58)}$$

Both treated and effluent discharged to surface water, $HUM_{effluent}(i,t)$ and $HUM_{c,untreat}(i,t)$, is distributed not just by population, but by a function of the urban population, prioritizing areas where sewer networks are concentrated. We define a per-cell "urban-ness," $urb(m,t)$, as a logistic function of gridded population density $POP(m,t)$, centered on an urban-cluster threshold, POP_{thresh} , of 300 persons km^{-2}[100] and used here as a continuous urban-weighting proxy:

$$urb(m,t) = \frac{1}{(1+exp[-k(log_{10}POP(m,t)-log_{10}POP_{thresh})]} \quad \text{(Eq. 59)}$$

$$\phi_{HUM_{effluent}} = POP(m,t) * urb(m,t) \quad \text{(Eq. 60)}$$

where k=1.5, so that $urb(m,t)$ values approach 1 in densely populated cells and approach zero in sparsely populated cells.

The two cropland-recycling streams, $HUM_{crop,nc}(i,t)$ and $HUM_{crop,bio}(i,t)$, are also split differently, reflecting management-based differences in distribution. Non-sewered excreta recycling (night soil) is typically applied to cropland close to where excreta is generated, both by rural households recycling locally and by transport of collected excreta from population centers to nearby farmland. It is therefore distributed across cropland without an urban or rural bias, on the same CROPGRIDS physical cropland area used for livestock manure:

$$\phi_{HUM_{crop,nc}} = H(m,t) \quad \text{(Eq. 61)}$$

In contrast, sewered biosolids, generated at urban treatment plants and applied to nearby farmland, follow peri-urban cropland and thus are spatially distributed in SCALES-global closer to urban areas:

$$\phi_{HUM_{crop,bio}} = H(m,t) * urb(m,t) \quad \text{(Eq. 62)}$$

Landfilled-sludge and non-collected human waste that reach the non-crop landscape are, like food waste, spatially distributed according to population alone.

$$\phi_{HUM_{landfill}} = \phi_{HUM_{terr,nc}} = POP(m,t) \quad \text{(Eq. 63)}$$

Surplus Calculations

From the budget terms above we compute three nitrogen surplus products at both country and gridded scales. These products differ in both the inputs and outputs included and in the land area over which the surplus is expressed (**Figure 5**). Each is a complete balance on its own land base, so the three are not additive.

The **cropland N surplus** is the agronomic surplus of cropland, expressed per hectare of cropland and directly comparable to the FAO Cropland Nutrient Budget:

$$NS_{cropland} = \frac{FERT_{crop,net}+BNF_{crop,leg}+BNF_{crop,free}+DEP_{crop}+MAN_{crop,net}+HUM_{crop,nc}+HUM_{crop,bio}-CROP}{AREA_{crop}(i,t)} \quad \text{(Eq. 64)}$$

The **landscape N surplus** extends the balance to the entire land surface, but excludes human waste in wastewater treatment plant effluent. It includes fertilizer applied to grazing land, deposition over all land, and pasture-deposited manure, all per hectare of land area

$$NS_{landscape} = \frac{FERT+BNF+DEP+MAN+HUM_{crop}+HUM_{landfill}+HUM_{terr,nc}+FOOD_{waste}-CROP-FORAGE}{AREA_{land}} \quad \text{(Eq. 65)}$$

The **total N surplus** adds N from untreated human waste and treated effluent to the landscape surplus, giving the complete anthropogenic N surplus of the land surface.

$$NS_{total} = NS_{landscape} + \frac{HUM_{N,effluent}+HUM_{c,untreat}}{AREA_{land}} \quad \text{(Eq. 66)}$$

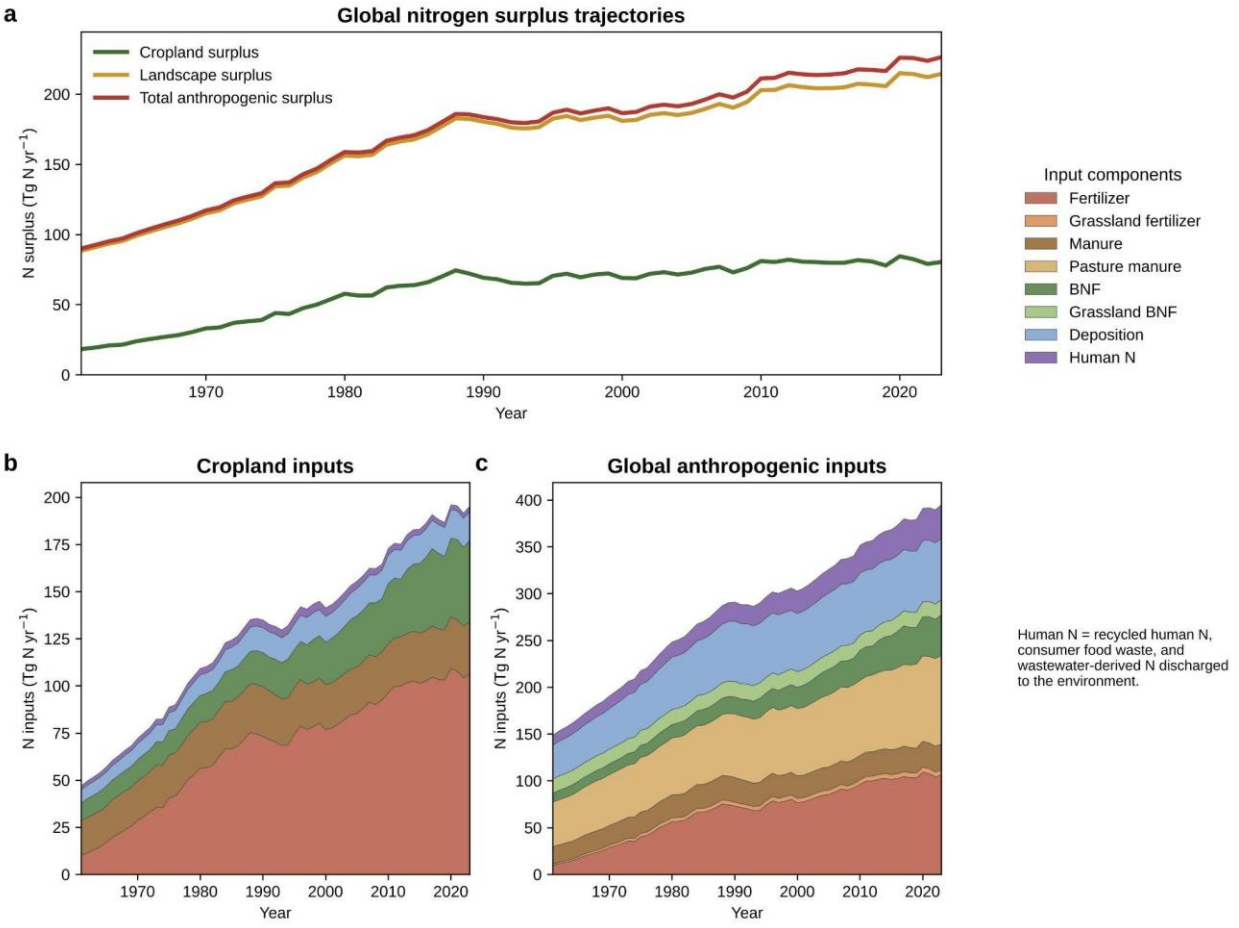


Figure 5. Evolution of the global nitrogen budget represented by SCALES-Global, 1961–2023. (a) Global nitrogen surplus trajectories for the three SCALES-Global accounting views: cropland surplus, landscape surplus, and total anthropogenic surplus. (b) Major N inputs contributing to the cropland surplus, including synthetic fertilizer, livestock manure applied to cropland, biological N fixation, atmospheric deposition, and recycled human nitrogen returned to cropland. (c) Major nitrogen inputs contributing to the total anthropogenic nitrogen budget, including agricultural inputs, atmospheric deposition, and aggregated human N inputs, including N discharged from wastewater treatment plants. Values represent annual global totals (Tg N yr^{-1}).

Uncertainty Analysis

Each budget term is subject to two types of uncertainty: (1) coefficient uncertainty and (2) activity (input) data uncertainty. For the coefficient uncertainty, each coefficient is assigned a coefficient of variation (CV) based on published uncertainty estimates or intercomparison studies representing error in the globally or regionally applied coefficients, e.g. crop N content, biological N fixation, livestock excretion rates, volatilization fractions, and sanitation parameters (**Table 8**). Activity data, which include fertilizer use, crop production, livestock inventories, grazing-land area and population, are assigned uncertainty based on the reliability of the underlying FAOSTAT reporting domain (**Table 9**).

Every national, global, and gridded SCALES-Global estimate is accompanied by an uncertainty estimate derived from Monte Carlo propagation through the complete nitrogen mass balance, using 10,000 Latin hypercube realizations.[101] In each realization, uncertain coefficients and activity data were sampled from

probability distributions defined from the literature.[7,102] Latin hypercube sampling was used to efficiently sample the multidimensional parameter space while maintaining stable estimates of the output distributions.

Assignment of Uncertainty

In each realization of the model runs, every coefficient is sampled from its own distribution and all budget terms are recomputed from the sampled values, so that a coefficient appearing in more than one term (e.g. $c_N(j)$ in Eqs. 21 and 27) takes the same value in each, and partitioned quantities (e.g. fertilizer split by *cf(i)*, Eqs. 3–4, and excreta split by *c(i,t)*, Eqs. 38–39) continue to sum to their totals. Sampled fractions are bounded at one, and the sludge-routing fractions are renormalized to sum to one. Livestock excretion rates are applied within FAOSTAT and are therefore represented by a single regional factor applied to both cropland and grazing-land manure.

Table 8. Coefficients sampled in the Monte Carlo analysis and their literature sources. All factors are assumed to have a log-normal distribution.

Coefficient	Eq.	Terms affected	CV	Basis for the CV
Fertilizer				
$cf(i)$	3–4	$FERT_{crop,appl}$, $FERT_{graze,appl}$	0.11	Based on uncertainty in allocations between crops and pasture (FAO Cropland Nutrient Budget)[35]
$fracGASF(p)$	5–7	Fertilizer NH_3 loss	0.70	Volatilization loss uncertainty based on IPCC data.[41]
Crop removal and biological N fixation				
$cN(j)$	21, 27	$CROP$, $BNF_{crop,leg}$	0.18	Crop nitrogen content varies across space and time, and the range based on that.[44]
$f_{bnf}(j)$	21	$BNF_{crop,leg}$	0.20	Based on observations of fixation compiled by Peoples et al.[48]
$wpf(j)$	21	$BNF_{crop,leg}$	0.25	Based on the uncertainty in legume fixation estimated for the FAO Cropland Nutrient Budget.[35]
r_{free}	22	$BNF_{crop,free}$	0.70	Based on the uncertainty in free-living fixation estimated for the FAO Cropland Nutrient Budget.[35]

r_{graze}	23	BNF_{graze}	0.40	Assumed; no published uncertainty exists for an area-averaged rate of pasture-legume fixation.
Manure				
Excretion rates	12–13	$MAN_{crop,appl}$, $MAN_{graze,appl}$	0.48	Uncertainty in manure excretion rates from FAO nutrient budgets.[35]
$fracGASM_{crp}$	12	Manure NH_3 loss, cropland	0.37	IPCC estimates.[36,41]
$fracGASM_{graze}$	13	Manure NH_3 loss, grazing land	0.18	Reported range for grazing-deposited excreta.[37,38]
Deposition				
$W(i,t)$, $I(m,t)$	31–33	DEP_{crop}, $DEP_{non-crop}$	0.15-0.35, by region	Based on the spread among the four deposition products of Vishwakarma et al.[13]
Forage				
$p_{managed}$	29	*FORAGE*	0.25	Based on measured N yields of managed grassland, 100–300 kg N ha^{-1} yr^{-1}.[57,58]
p_{range}	29	*FORAGE*	0.33	Measured N yields of rangeland, 10–50 kg N ha^{-1} yr^{-1}.[57,58]
Human-derived N				
Protein N content	34	All human terms	0.05	Range of protein-to-N conversion factors across foods.[64,103]
$w_C(i,t)$	35–36	$FOOD_{waste,N}$ and, through the complement, all excreta terms	0.30, by region	Uncertainty of approximately ±30% in retail and household loss fractions.[104]

$c(i,t)$	38–39	Sewered and non-sewered terms	0.20, by region	Error of the income-based sewerage model (Eqs. 40, 43) against observed connection in 100 countries
$f_{treat}(i,t)$	44–46	Treated and untreated sewage	0.15, by region	Error of the regional treated-to-collected ratios (Table 6) against observed values in 101 countries
R_{low}	47	HUM_{den}, $HUM_{effluent}$, HUM_{sludge}	0.40	Van Drecht et al.[19]
R_{high}	47	HUM_{den}, $HUM_{effluent}$, HUM_{sludge}	0.10	Based on N balances of activated-sludge plants with denitrification.[72–74]
sl_{bio}	48–50	HUM_{sludge}, HUM_{den}	0.20	Based on sludge fractions in closed plant N balances.[72–74]
$a_{ag}(i,t)$, $a_{lf}(i,t)$	51–53	$HUM_{crop,bio}$, $HUM_{landfill}$, HUM_{inc}	0.35, by region	Assumed; the routing shares are compiled from national statistics of uneven quality and interpolated between anchor years (Supplemental Table S2)
$\lambda_{ns}(i,t)$	54–57	$HUM_{crop,nc}$, $HUM_{terr,nc}$	0.40, by region	Corresponds to assignment one class away in the recycling classes of Morée et al..[20]

Next, activity-data uncertainty is assigned according to the reliability of the underlying FAOSTAT reporting domain (e.g., crop production, fertilizer consumption, livestock inventories, and population) (**Table 9**). The activity-data CV values in the table are derived from the provenance flag that FAOSTAT attaches to each reported value, so that reporting uncertainty is specific to each national input rather than uniform. Each value is flagged as an official figure (A), a figure supplied by an international organization (X), an FAO estimate (E), or an imputed gap-fill (I), which together form a reliability ordering from reported to modeled data. We assign each flag a multiplier of a term-specific baseline CV, increasing from official to imputed (1.0, 1.2, 1.6, and 2.4, respectively, and 3.0 for the rare missing-and-filled values), and combine the flags within each country and year as a value-weighted mean, so that a country's activity CV reflects the reliability of its largest flows. Each term's baseline is set so that its value-weighted global mean equals the term's assigned global activity CV, so the flags redistribute the reporting uncertainty across countries and years without altering its global magnitude. Baseline magnitudes are drawn from the IPCC[105] and the FAOSTAT cropland nutrient budget.[35] Provenance flags are taken from the FAOSTAT domain supplying each term's activity data: fertilizer use for synthetic fertilizer, crop production for crop N removal and biological N fixation, reported livestock inventories for

manure, permanent-pasture area for grazing terms, and dietary protein supply for human N. Note that atmospheric deposition, a modeled field, carries no activity channel. The resulting activity CV varies by country and year and is largest where inputs are estimated or imputed rather than officially reported. This flag-based scheme follows the country-tiered treatment of activity-data uncertainty in the EDGAR emissions inventory.[106] All sampled coefficients and activity factors are assumed to follow lognormal distributions.

Table 9. Activity (reporting) coefficient of variation (CV) values applied to each budget term according to the FAOSTAT provenance flag of each national value. Flag designations are as follows: A, official figure; X, international organization; E, estimated; I, imputed; M, missing. A is the official baseline. Manure provenance follows reported livestock head counts and the human term follows dietary protein supply; deposition has no activity channel. Dashes denote flags that do not occur for a term.

Term	A (official)	X	E	I	M
Fertilizer	0.102	0.122	0.163	0.244	0.305
Crop production	0.088	0.106	0.141	0.212	0.265
BNF	0.088	0.106	0.141	0.212	0.265
Forage / grazing	0.054	0.064	0.086	0.129	0.161
Manure (livestock head counts)	0.064	0.077	0.103	0.154	0.192
Human (protein supply)	0.031	0.038	0.050	0.075	0.094

The two sources of uncertainty (coefficient and activity data) are sampled independently. Coefficient uncertainties are applied at the spatial scale at which each coefficient is defined. Coefficients taken as a single global value, e.g., crop N content, biological N fixation coefficients, and volatilization fractions, are drawn once per model run and applied identically to every country (common-mode uncertainty). Coefficients defined by region, e.g. livestock excretion rates, which follow the IPCC world regions, together with atmospheric deposition and the sanitation coefficients, are drawn once per region per model run. In contrast, activity-data uncertainty is treated as independent across countries based on FAOSTAT data reporting.[105]

Gridded uncertainty

At the grid scale, each cell holds a share of its country's national total for each component of the mass balance. Because each cell is a fixed share of a total whose uncertainty spread we have already characterized nationally, we propagate the national Monte Carlo uncertainty of each component to each cell via closed-form error propagation rather than running a separate Monte Carlo for every cell.[107] The variance, $\sigma^2_{surplus}(m)$, of a cell's surplus is therefore calculated as follows:

$$\sigma^2_{surplus}(m) = \sum_k \quad \sum_l \quad s_k s_l x_k(m) x_l(m) C_{kl}(i) \qquad \text{Eq. 67}$$

where $x_k\ (m)$ is the value of the budget component *k* in cell *m* of country *i*, s_k is +1 for inputs and -1 for outputs, and $C_{kl}\,(i)$ is the relative covariance of components k and l across the national Monte Carlo realizations for country *i*, so that C_{kk}(i) is the squared CV of component *k*. The off-diagonal terms carry the dependence between components that share a coefficient or a partition.

Scope of the uncertainty estimates

The reported uncertainty intervals quantify uncertainty in model parameters and input activity data conditional on the adopted accounting framework. They do not include structural uncertainty arising from alternative representations of biological fixation, manure management, sanitation pathways, or spatial downscaling, and uncertainties in the spatial proxy layers used for downscaling are not propagated. The reported intervals should therefore be interpreted as parametric uncertainty within the SCALES-Global framework rather than encompassing all possible model structures.

For the 2023 reconstruction (**Figure 6**), the Monte Carlo mean closely matches the deterministic central estimate for all three accounting views (within 1%), confirming that the uncertainty propagation preserves the underlying mass balance without introducing systematic bias. The 95% intervals are 37-119 Tg N yr^{-1} for the cropland surplus (central estimate 80 Tg N yr^{-1}), 164-274 Tg N yr^{-1} for the landscape surplus (214 Tg N yr^{-1}), and 176-286 Tg N yr^{-1} for the total anthropogenic surplus (226 Tg N yr^{-1}), corresponding to coefficients of variation of 26%, 13% and 12%, respectively. Relative uncertainty decreases as the accounting domain expands, with the cropland surplus exhibiting the largest uncertainty and the total anthropogenic surplus the smallest. This behavior is expected because the cropland surplus is calculated as the residual of two large opposing fluxes (inputs and crop removal), whereas the landscape and total surpluses are larger integrated quantities for which equivalent absolute uncertainties represent a smaller fraction of the total. Over 1961–2023, the CV ranged from 15% to 32% for the cropland surplus and from 9-13% for the landscape and total surpluses. At the grid scale, the median coefficient of variation in 2023, over cells with an absolute surplus above 5 kg N ha^{-1}, was 0.46 for the cropland surplus and 0.30 for the landscape and total surpluses. Representative uncertainty ranges for the three accounting views are shown in Fig. 6.

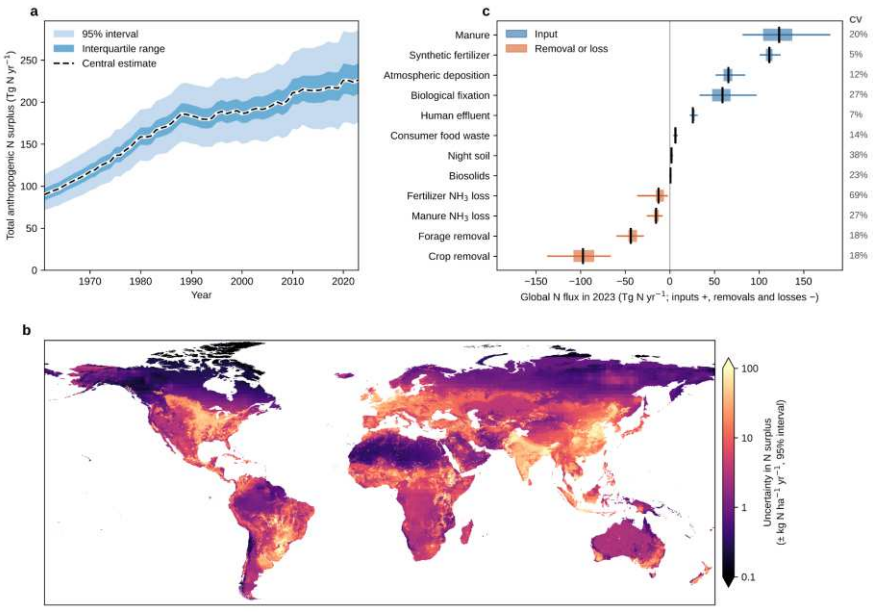


Figure 6. Monte Carlo uncertainty of the SCALES-Global nitrogen budget. Uncertainty is quantified from 10,000 Monte Carlo realizations in which the coefficients and activity data are sampled from their probability distributions (Tables 8 and 9) and propagated through the complete nitrogen mass balance. (a) Global total anthropogenic nitrogen surplus (1961–2023) showing the deterministic estimate (dashed line), the Monte Carlo interquartile range (dark shading), and 95% interval (light shading). (b)

Uncertainty in the 2023 gridded total anthropogenic nitrogen surplus at 0.05° resolution, expressed as the amount by which the surplus in each cell may differ from its central estimate at the 95% level (± kg N ha^{-1} yr^{-1}, logarithmic scale). (c) Global nitrogen-budget components for 2023, with inputs in blue and removals or losses in orange. Black lines mark the central estimates, boxes the interquartile ranges, and horizontal lines the 95% intervals; the right-hand column gives the coefficient of variation of each component in the Monte Carlo analysis.

Potential use

The dataset supports global/regional environmental monitoring, N-hotspot identification, and parameterization of biogeochemical and water-quality models. The 0.05° resolution enables finer-scale input/output estimation than country-scale budgets; the 1961–2023 coverage supports legacy-N and long-term trend analyses. Mass-conserving downscaling means the grid can be aggregated to any administrative or watershed boundary while remaining consistent with the validated national budget, and the multiple surplus calculations (cropland/land/total) let users select the accounting scope appropriate to their application.

Data Records

The SCALES-Global dataset is publicly available through Zenodo, https://doi.org/10.5281/zenodo.23670209; all versions, https://doi.org/10.5281/zenodo.22737030. It consists of three complementary products that provide annual nitrogen budgets from 1961-2023 at gridded, national, and global scales.

Gridded annual N budget

Annual gridded products are provided as CF-compliant NetCDF-4 files, one per variable, each holding all years from 1961 to 2023 (full/<variable>_1961_2023.nc; dimensions time × lat × lon). A smaller lite/ set holds the three surplus views, cropland area and cell area for mapping. Because Zenodo stores files in a flat list, the folder is written as a file-name prefix (e.g. full__fert_kg_1961_2023.nc).

Spatial characteristics:

- Geographic coordinate system: WGS84 (EPSG:4326)
- Resolution: 0.05° (~5.5 km)
- Grid dimensions: 7200 × 3600 cells
- Latitude ordered south-to-north

The variables are listed in **Table 10**. All component variables are reported as kg-N $cell^{-1}$ yr^{-1}, while surplus products are reported as kg-N ha^{-1} yr^{-1} on their respective land bases.

952 **Table 10.** Variables in the gridded dataset (one NetCDF file per variable).

953

Variable	Description	Units
fert_kg, fert_grass_kg, bnf_kg, bnf_grass_kg, dep_kg, dep_land_kg, man_crop_kg, man_past_kg, food_waste_kg, hum_crop_kg, hum_ns_kg, hum_bio_kg, hum_sew_land_kg, hum_sew_water_kg	Per-cell N inputs: synthetic fertilizer (cropland, grazing land), biological fixation (cropland, grazing land), deposition (cropland, all land), manure (applied to cropland, deposited on grazing land), consumer food waste, and human N (to cropland, night soil, biosolids, non-collected excreta to land, sewered discharge to surface water)	kg-N $cell^{-1}$ yr^{-1}
crop_kg, forage_kg	Per-cell N removal in harvested crops and in forage from grazing land	kg-N $cell^{-1}$ yr^{-1}
fert_vol_kg, man_vol_past_kg	Per-cell NH_3 volatilization from fertilizer and from grazing-land manure	kg-N $cell^{-1}$ yr^{-1}
cropland_ha	cropland area per cell	ha
surplus_cropland_kgha	View 1: cropland surplus	kg-N ha^{-1} cropland yr^{-1}
surplus_landscape_kgha	View 2: landscape surplus	kg-N ha^{-1} land yr^{-1}
surplus_full_kgha	View 3: total anthropogenic surplus	kg-N ha^{-1} land yr^{-1}
cropland_frac	Cropland areas as a fraction of the cell area (HYDE 3.3)	fraction (0-1)

954 **National annual nitrogen budgets**

955 Annual country-scale budgets are provided as a single comma-separated table,
956 tables/national_budget_N_1961_2023.csv, with one row per country and year; the same data are
957 provided as one file per year in tables/per_year_national_N_1961_2023.zip.

958 The table contains one record for every country and year. Component terms are reported both as total
959 annual mass (kg-N yr^{-1}) and normalized by the appropriate land area (kg-N ha^{-1} yr^{-1}).

960 Variables include fertilizer, biological N fixation, atmospheric deposition, livestock manure, human-
961 derived N, crop removal, forage removal, and surplus (cropland, landscape, and total anthropogenic).

Global annual time series

Global summaries are provided in tables/global_timeseries_N_1961_2023.csv. The file contains annual global totals (1961-2023) for every component of the nitrogen budget, including fertilizer, biological N fixation, deposition, manure, human-derived nitrogen, crop removal, forage removal, and the three surplus products. The file also reports global cropland area, total land area, and surplus intensities (kg-N ha^{-1} yr^{-1}).

Uncertainty estimates

National and global uncertainty are included in the two tables above: for each country and year, and for the global total, for each of the three surplus views, the central estimate, Monte Carlo mean, standard deviation, 2.5th and 97.5th percentiles, and coefficient of variation. Gridded uncertainty is provided in the uncertainty/ folder for 1961, 1965, 1970, 1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015, 2020, and 2023 on the same grid as the budget files; each surplus view contains the standard deviation, its coefficient and activity-data parts, and the half-width of the 95% interval, in kg-N ha^{-1} yr^{-1}.

Technical Validation

We assess SCALES-Global in four complementary ways: (i) an internal-consistency check that confirms the country-scale budget correctly reproduces the reference budget it is built to match; (ii) comparison against independent datasets that were built from different data and methods; (iii) component comparisons for gridded fertilizer and for wastewater nitrogen; and (iv) checks of numerical integrity and temporal plausibility (**Table 11**).

Internal consistency with the FAO Cropland Nutrient Budget (country scale, 2020)

Because SCALES-Global adopts the same fertilizer, manure, crop-removal, and coefficient framework as the FAO Cropland Nutrient Budget,[35] close agreement is expected with our national-scale analysis. The comparison therefore serves as an implementation check rather than an independent validation. Across countries, the individual budget components are in close agreement (fertilizer $R^2 = 0.97$, manure $R^2 = 0.98$, biological N fixation $R^2 = 0.99$, deposition $R^2 = 0.96$, and crop removal $R^2 = 0.77$), and the integrated cropland surplus agrees at $R^2 = 0.83$ ($p<0.001$, **Fig. 7a)**, with the SCALES-Global total approximately 8% lower than the FAO total. These results demonstrate that the SCALES-Global implementation faithfully reproduces the reference national accounting framework on which it is based.

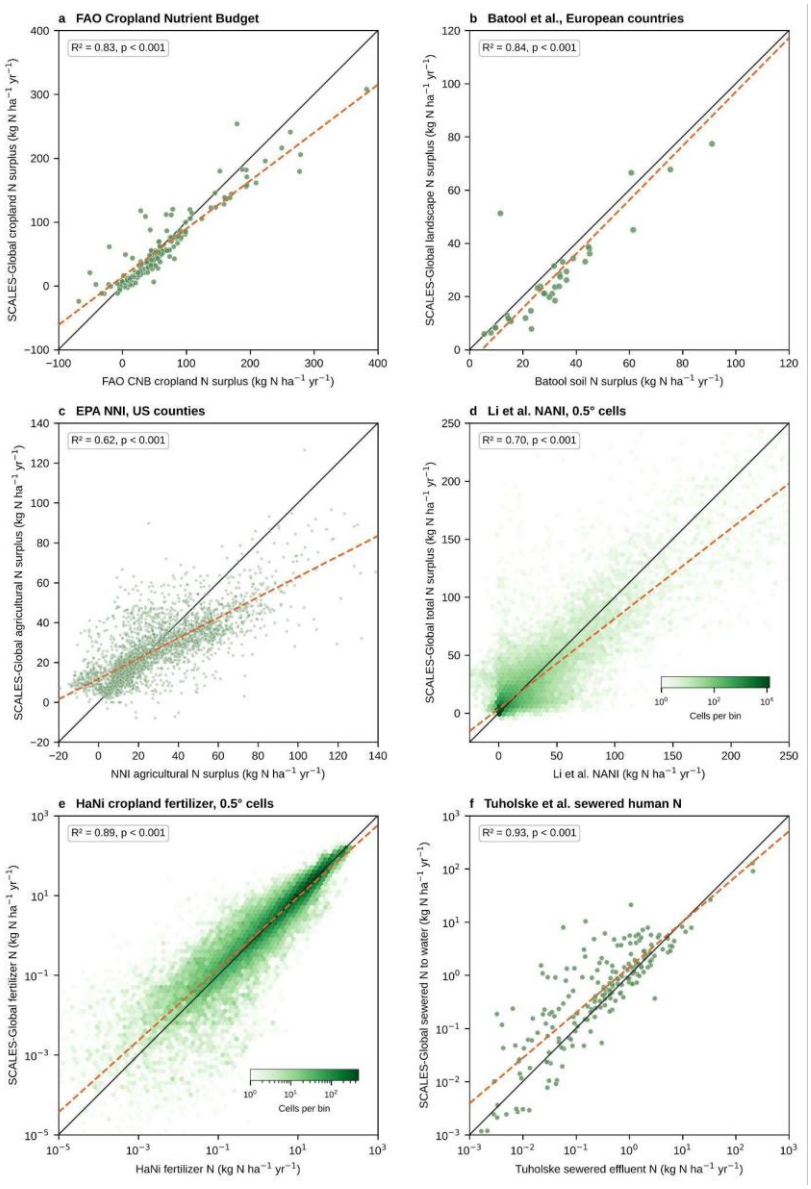


Figure 7. Validation of SCALES-Global through reproduction of the reference FAO budget, comparison with independent benchmark datasets and comparison of individual budget components. The figure panels show comparisons of SCALES-Global against **(a)** the 2020 FAO Cropland Nutrient Budget[35] (country scale), **(b)** the European soil nitrogen surplus dataset (2019),[15] **(c)** the county-scale US National Nutrient Inventory (2017),[108] **(d)** the global Net Anthropogenic Nitrogen Inputs (NANI) dataset (2015),[109] **(e)** the HaNi global fertilizer dataset (2019), [14] and **(f)** the Tuholske et al. sewered human N dataset (2015).[110]

The residual scatter and the lower SCALES-Global total arise from three intentional differences in budget construction rather than implementation error. First, SCALES-Global subtracts post-application NH_3 volatilization losses from synthetic fertilizer and manure before calculating surplus, whereas the FAO-CNB is a gross soil-surface balance that does not include this correction. Second, manure is partitioned between cropland and grazing land using the FAOSTAT EMN livestock-management framework, resulting in manure inputs to cropland being approximately 7% greater than the aggregate “manure

applied to soils" values used by the FAO-CNB. Third, SCALES-Global corrects eight crop N removal coefficients in the Ludemann et al. world coefficient set used by the FAO-CNB, as described in the Methods (Crop N removal, Eq. 27). Applied to 2020 production, these corrections lower global crop N removal by approximately 8 Tg N yr^{-1} and account for nearly all of the difference in crop removal. With the original coefficients, the two crop-removal estimates agree at $R^2 = 1.00$ and the cropland surplus at $R^2 = 0.92$. Together, these methodological differences explain the deviations between the two datasets while confirming that SCALES-Global can accurately reproduce the FAO-CNB accounting framework.

Table 11. Technical validation of the SCALES-Global nitrogen-surplus products. All correlations are significant at p<0.001.

Evaluation category	Reference / Test	Scale, Year	SCALES-Global product evaluated	Key result
Internal Consistency Check	FAO Cropland Nutrient Budget[35]	Country, 2020	Cropland surplus	$R^2 = 0.83$ (p<0.001), Fig. 7a
External Comparison	Batool et al. N Surplus [15]	Country (Europe), 2019	Landscape surplus	$R^2 = 0.84$ (p<0.001)
External Comparison	EPA Next-Generation National Nutrient Inventory[108,111]	County (US), 2017	Agricultural surplus (six components)	$R^2 = 0.62$ (p<0.001)
External Comparison	Li et al. NANI[109]	Global 0.5° grid, 2015	Total anthropogenic surplus	$R^2 = 0.70$ (p<0.001)
External Comparison	Tian et al. HaNi[14]	Global 0.5° grid, 2019	Cropland fertilizer	$R^2 = 0.89$ (p<0.001)
External Comparison	Tuholske et al. Sewered N discharge[110]	Country, 2015	Sewered human N discharged to surface water	$R^2 = 0.93$ (p<0.001)
Numerical integrity Check	Internal conservation of mass	Country, 1961, 1980, 2000, 2020, 2023	All gridded budget terms	$R^2 = 1.00$ ($p < 0.001$); 99.6% of national totals reproduced within ±1%

Comparison with external datasets

SCALES-Global was evaluated against independently developed datasets spanning national, subnational, and global scales. Because the reference products differ in accounting boundaries, land bases, spatial resolution, and underlying data, each comparison uses the SCALES-Global accounting view and budget component most closely aligned with the corresponding benchmark. These comparisons therefore assess complementary aspects of the dataset rather than a single common quantity.

Europe

Batool et al. (2022)[15] report country-scale soil nitrogen surpluses for European countries that include both cropland and pasture systems. We therefore compare their 2019 estimates with the SCALES-Global landscape surplus, excluding human waste-derived N, as the Batool et al. budget does not include these

inputs. The surplus compared includes fertilizer applied to cropland and grazing land, atmospheric N deposition over all land, and manure applied to cropland and deposited on grazing land, and crop and forage N removal, expressed per hectare of total land, which is the matching scope. Across 37 European countries the agreement is strong (R^2 = 0.84, p<0.001; **Fig. 7b**). Differences arise from contrasts in fertilizer and manure reconstructions, treatment of grazing land N removal, atmospheric deposition, and the land area used for normalization. Because the two products were developed independently and use different historical data streams, available at different spatial scales, the comparison provides an external country-scale evaluation of the landscape accounting view.

United States county-scale surplus

The EPA Next-Generation National Nutrient Inventory (NNI)[108,111] provides county-scale net anthropogenic nitrogen fluxes for the contiguous United States (1950/1987–2017). The 0.05° SCALES-Global grid was aggregated to county boundaries and compared with six budget components for the year 2017: fertilizer, manure, biological N fixation, atmospheric deposition, crop removal, and forage removal. The agricultural surplus built from these six components agrees across 3,108 US counties at R^2 = 0.62 (p<0.001; **Fig. 7c)** and the SCALES-Global total is within 7% of the inventory total. Among the individual components, agreement is strongest for crop removal (R^2 = 0.84), manure (R^2 = 0.84), and biological N fixation (R^2 = 0.82), followed by fertilizer (R^2 = 0.60), and weakest for atmospheric deposition (R^2 = 0.28) and forage removal (R^2 = 0.18). Manure magnitudes run approximately 19% below the NNI values because SCALES-Global takes into account the management and volatilization losses of N from manure, whereas the NNI reports gross excreted manure. Grazing-land forage removal is the least constrained of the six fluxes: the SCALES-Global total is roughly half of the NNI value, and the county-scale agreement is weak. Deposition agrees well in overall magnitude, but less well with regard to spatial patterns. This difference is primarily due to different sources of deposition data. The NNI reports total wet and dry deposition of oxidized and reduced N over each county, derived from US deposition maps: the NADP Total Deposition (TDep) product, which fuses monitoring-network measurements with CMAQ simulations, for 2000–2017, and CMAQ simulations for 1990–2010, reconciled by spatial regression.[108,112] The TDep data is not a global product, making its use inappropriate for use in SCALES-Global, which instead takes the national deposition magnitude from the global Vishwakarma et al. product and resolves within-country differences only through the 0.5° ISIMIP3a gradient, which cannot reproduce gradients at the finer scale of the TDep product. This comparison with the EPA NNI provides an independent, sub-national evaluation in a data-rich region and indicates that the global framework reproduces broad county-scale variation in the anthropogenic N surplus.

Global spatial pattern of anthropogenic nitrogen inputs

Li et al. (2024)[109] map global net anthropogenic nitrogen inputs (NANI) at 5-arcmin resolution (~9 km) for the year 2015. While there are some differences between the NANI product and our N surplus products, their dataset is comparable in scope with the SCALES-Global total anthropogenic surplus, and it provides an independent representation of its spatial distribution. We therefore compare the Li et al.[109] product with the SCALES-Global total anthropogenic surplus after aggregating both datasets to a common 0.5° grid. Across 67,273 land cells, the products show strong spatial correspondence (R^2 = 0.70, p < 0.001; **Fig. 7d**), with the SCALES-Global total over the matched cells being approximately 9% higher (214 versus 197 Tg N). Differences in both slope and magnitude are expected because the two accounting frameworks treat livestock and human waste differently. NANI estimates N from livestock and human waste through the net food and feed balance,[18] whereas SCALES-Global represents a land-surface surplus that explicitly subtracts forage removal and NH_3 volatilization. In addition, the NANI

framework includes only oxidized N deposition, whereas SCALES-Global includes both oxidized *and* reduced N in the budget. Finally, SCALES-Global routes human waste through separate recycling and wastewater pathways. This comparison therefore evaluates broad spatial correspondence rather than equality of the two accounting quantities.

Gridded fertilizer inputs

HaNi[14] provides annual gridded (5 arcmin, ~9-km) nitrogen inputs by land use (1860–2019). Because HaNi reports inputs rather than surplus magnitudes, we use it to evaluate the spatial allocation of individual SCALES-Global components. Our gridded cropland fertilizer agrees strongly with HaNi in both pattern and magnitude ($R^2 = 0.89$, $p < 0.001$; n = 38,024 cells; **Fig. 7e**), with matched totals of approximately 98 Tg-N in HaNi and 103 Tg-N in SCALES-Global. Comparisons for 1980 and 2000 produced a similarly strong correspondence.

Wastewater nitrogen

Tuholske et al.[110]provide an independent, spatially explicit global estimate of human sewage N for the year 2015. Their estimate is derived from national protein supply and WHO/UNICEF sanitation statistics. We compare their values for national sewered effluent N with the SCALES-Global sewered N discharged to surface water values for 2015, per hectare of land. Across 193 countries the two products agree at $R^2 = 0.93$ ($p < 0.001$; **Fig. 7f**). The differences in magnitude arise mainly from the treatment of collected sewage. Tuholske et al. assume a uniform 55% removal of N from all sewered human waste, whereas SCALES-Global distinguishes between treated and non-treated sewered waste. More specifically, untreated sewage is assumed to be discharged in full to surface water; in contrast, N is removed from the treated share, at an efficiency that increases with GDP from 10% to 80%, following the primary- and tertiary-treatment values of Van Drecht et al. and Moree et al (Eq. 47).[19,20] SCALES-Global discharge estimates are therefore higher than the Tuholske et al. estimates in countries where a large fraction of the sewage is untreated, including China (71% of the population connected, 51% treated) and India (35% connected, 11% treated), and lower than the Tuholske et al. estimates in high-income countries with advanced treatment such as the United States and Germany. Globally, the effect of untreated dominates, and the SCALES-Global sewered discharge is therefore 35% higher than the Tuholske estimate (9.7 versus 7.2 Tg N). Despite these differences in magnitude that follow from an identifiable treatment assumption, the close agreement in spatial pattern between SCALES-Global and Tuholske et al. supports the SCALES-Global representation of wastewater N.

Numerical integrity and temporal plausibility

Mass conservation

In addition to comparisons with independent datasets, we evaluated the numerical integrity of the spatial downscaling procedure. Because every gridded budget component is generated by distributing nationally estimated fluxes using normalized spatial weights (Eqs. 1-2), the sum of all grid-cell values within a country should equal the corresponding national total. Across all countries and budget terms in five different years (1961, 1980, 2000, 2020, and 2023, n=5,826 country-term pairs), the median ratio of gridded to national totals was 1.000, with 99.6% of country totals being reproduced within ±1%. The small number of countries exhibiting slightly lower ratios (typically 0.97–0.99) occur where a few coastal cells fall outside the common land mask used by the gridded proxy layers, representing a minor edge effect during spatial allocation rather than a violation of mass balance. This evaluation confirms that the downscaling procedure conserves mass while preserving the independently calculated national budgets.

Whole-record consistency

Although external validation is necessarily limited to benchmark years and available comparison datasets, SCALES-Global is computed annually from 1961 through 2023 using a consistent accounting framework. The resulting global time series reproduces the well-documented evolution of anthropogenic nitrogen use associated with the intensification of global agriculture. Synthetic fertilizer inputs increase from approximately 10 Tg-N yr^{-1} in 1961 to approximately 106 Tg-N yr^{-1} in 2023, driving corresponding increases in cropland nitrogen surplus (approximately 18 to 80 Tg N yr^{-1}) and total anthropogenic nitrogen surplus (approximately 90 to 226 Tg N yr^{-1}). Over the full period, the global cropland surplus per hectare closely follows the FAO Cropland Nutrient Budget world series ($R^2 = 0.99$). These long-term trajectories are consistent with FAOSTAT fertilizer statistics and previously published global nitrogen-budget reconstructions, providing confidence that the dataset captures the temporal evolution of anthropogenic nitrogen loading over the full period of record rather than only the benchmark years used for external validation.

Summary

The SCALES-Global dataset is technically robust across multiple spatial scales, budget components, and independent reference datasets. The model faithfully reproduces the FAO Cropland Nutrient Budget accounting framework that forms its methodological foundation while also showing strong agreement with independent national, subnational, and global nitrogen-budget products, as well as the independently developed HaNi gridded fertilizer dataset. Component-specific comparisons further show that the gridded fertilizer allocation and the wastewater nitrogen pathway agree with independent global estimates. Finally, the spatial downscaling conserves mass, and the complete 1961–2023 reconstruction reproduces the well-established historical evolution of global anthropogenic nitrogen use. Together, these complementary evaluations provide confidence in both the national accounting framework and the gridded SCALES-Global products for analyses spanning local to global scales.

Usage Notes

SCALES-Global provides three complementary nitrogen surplus products derived from a common underlying mass balance. Users should select the accounting view that best matches their application. The cropland surplus (surplus_cropland_kgha) is directly comparable to the FAO Cropland Nutrient Budget and is appropriate for studies of agronomic nitrogen management, fertilizer-use efficiency, crop production, and agricultural sustainability. The landscape surplus (surplus_landscape_kgha) extends the accounting to the entire terrestrial landscape by incorporating grazing land fertilizer, pasture-deposited manure, forage removal, and terrestrial human nitrogen pathways, while excluding wastewater effluent discharged directly to receiving waters. It is intended for studies of terrestrial nitrogen accumulation and landscape-scale nutrient budgets. The total anthropogenic surplus (surplus_full_kgha) additionally includes wastewater-derived nitrogen discharged to surface waters and is the recommended product for studies of watershed nitrogen loading, riverine export, coastal eutrophication, and comparisons with Net Anthropogenic Nitrogen Input (NANI)-type frameworks.

Users should also consider several important limitations. First, historical cropland extent, grazing-land extent, and population distributions are reconstructed from HYDE 3.3, while crop-specific spatial distributions are based on the modern CROPGRIDS dataset. Consequently, although land use, fertilizer application, atmospheric deposition, and other budget terms evolve annually from 1961–2023, the within-cropland spatial distribution of individual crop types reflects the modern agricultural landscape.

Second, the allocation of manure between cropland and grazing land follows the FAOSTAT Emissions database and associated IPCC manure-management framework. This convention differs from that adopted by some other global products, including HaNi, and may contribute to differences in cropland versus grazing land manure inputs. Third, human-derived nitrogen pathways are reconstructed from historical sanitation, wastewater treatment, and biosolids management datasets that remain sparsely documented in many regions. Although these pathways are constrained by available national statistics and literature, they constitute one of the more uncertain components of the global nitrogen budget. Finally, the spatial downscaling allocates national budget terms using the best available global proxy datasets. Consequently, the gridded products represent the most likely spatial distribution of each flux at 0.05° resolution but should not be interpreted as resolving sub-grid-scale management practices or field-level nitrogen applications.

Code Availability

All processing code (national budget pipeline, downscaling engine, validation, and figure generation) is available at https://github.com/blue-node-hydrologics/SCALES-Global and archived on Zenodo (https://doi.org/10.5281/zenodo.23769024). The pipeline is written in Python (xarray, rasterio, pandas, geopandas) and runs from the public input datasets listed in the repository README.

Data availability

All input datasets are publicly available. National activity and budget data, including fertilizer use, crop and livestock production, livestock manure, land use, food supply, population, and the Cropland Nutrient Budget are from FAOSTAT.[49] Crop N-removal coefficients are from Ludemann et al. (Dryad).[44] Cropland extent and historical population are from HYDE 3.3,[23] crop-specific harvested area from CROPGRIDS,[24] and grazing-land extent from Global Pasture Watch,[25] with MODIS MCD12C1 land cover (NASA LP DAAC)[29] as an auxiliary grassland base. The within-country distributions of synthetic fertilizer and manure are from Coello et al.[12] and Zhang et al.,[31] respectively. Atmospheric N deposition combines the national magnitudes from Vishwakarma et al. (2023)[13] with the gridded pattern from ISIMIP3a.[33] The present-day population distribution is from GHS-POP,[32] sewer connection and wastewater treatment data from the UN Statistics Division[67,69] and OECD series,[68] and per-capita GDP from the Maddison Project Database 2023.[66] Two datasets compiled for this study are provided as Supplementary Information: country-specific sewer-connection anchors (Supplementary Table S1) and national trajectories of biosolids disposal routes (Supplementary Table S2). The SCALES-Global output dataset is available from the Zenodo repository, https://doi.org/10.5281/zenodo.23670209; all versions, https://doi.org/10.5281/zenodo.22737030.

Acknowledgements

This research was supported in part by a National Science Foundation CAREER award to KVM (EAR-2239715). NBB acknowledges support from the Canada Research Chairs Program and the Natural Sciences and Engineering Research Council of Canada Alliance Program.

Competing Interests

None

References

1. Galloway, J. N. *et al.* Transformation of the nitrogen cycle: recent trends, questions, and potential solutions. *Science* **320**, 889–892 (2008).
2. Rockström, J. *et al.* Planetary boundaries: Exploring the safe operating space for humanity. *Ecol. Soc.* **14**, (2009).
3. Steffen, W. *et al.* Sustainability. Planetary boundaries: guiding human development on a changing planet. *Science* **347**, 1259855 (2015).
4. Fowler, D. *et al.* Effects of global change during the 21st century on the nitrogen cycle. *Atmos. Chem. Phys.* **15**, 13849–13893 (2015).
5. Galloway, J. N. *et al.* Nitrogen cycles: past, present, and future. *Biogeochemistry* **70**, 153–226 (2004).
6. Erisman, J. W. *et al.* Consequences of human modification of the global nitrogen cycle. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130116 (2013).
7. Oenema, O., Kros, H. & Vries, W. de. Approaches and uncertainties in nutrient budgets: implications for nutrient management and environmental policies. *Eur. J. Agron.* **20**, 3–16 (2003).
8. Zhang, X. *et al.* Managing nitrogen for sustainable development. *Nature* **528**, 51–59 (2015).
9. Lassaletta, L. *et al.* Food and feed trade as a driver in the global nitrogen cycle: 50-year trends. *Biogeochemistry* **118**, 225–241 (2014).
10. Byrnes, D. K., Van Meter, K. J. & Basu, N. B. Long-term shifts in U.S. nitrogen sources and sinks revealed by the new TREND-nitrogen data set (1930–2017). *Global Biogeochem. Cycles* **34**, (2020).
11. Lu, C. & Tian, H. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth Syst. Sci. Data* **9**, 181–192 (2017).
12. Coello, F. *et al.* Global crop-specific fertilization dataset from 1961-2019. *Sci. Data* **12**, 40 (2025).
13. Vishwakarma, S., Zhang, X., Dobermann, A., Heffer, P. & Zhou, F. Global nitrogen deposition inputs to cropland at national scale from 1961 to 2020. *Sci. Data* **10**, 488 (2023).

14. Tian, H. *et al.* HaNi: A Historical dataset of Anthropogenic Nitrogen Inputs to the terrestrial biosphere (1860-2019). PANGAEA https://doi.org/10.1594/PANGAEA.942069 (2022).
15. Batool, M. *et al.* Long-term annual soil nitrogen surplus across Europe (1850–2019). *Scientific Data* **9**, 1–22 (2022).
16. Gu, B., Ju, X., Chang, J., Ge, Y. & Vitousek, P. M. Integrated reactive nitrogen budgets and future trends in China. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8792–8797 (2015).
17. Chang, S. Y., Byrnes, D. K., Basu, N. B. & Van Meter, K. J. gTREND-Nitrogen - Long-term nitrogen mass balance data for the contiguous United States (1930-2017). *Sci. Data* 1–19 (2026) doi:10.1038/s41597-026-06576-x.
18. Hong, B., Swaney, D. P. & Howarth, R. W. A toolbox for calculating net anthropogenic nitrogen inputs (NANI). *Environ. Modell. Softw.* **26**, 623–633 (2011).
19. Van Drecht, G., Bouwman, A. F., Harrison, J. & Knoop, J. M. Global nitrogen and phosphate in urban wastewater for the period 1970 to 2050: N AND P IN URBAN WASTE WATER. *Global Biogeochem. Cycles* **23**, (2009).
20. Morée, A. L., Beusen, A. H. W., Bouwman, A. F. & Willems, W. J. Exploring global nitrogen and phosphorus flows in urban wastes during the twentieth century: TWENTIETH CENTURY URBAN N AND P FLOWS. *Global Biogeochem. Cycles* **27**, 836–846 (2013).
21. Eicher, C. L. & Brewer, C. A. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartogr. Geogr. Inf. Sci.* **28**, 125–138 (2001).
22. de Belém Costa Freitas Martins, M., de Sousa Xavier, A. M. & de Sousa Fragoso, R. M. Redistributing agricultural data by a dasymetric mapping methodology. *Agric. Resour. Econ. Rev.* **41**, 351–366 (2012).
23. Klein Goldewijk, K. HYDE 3.3: History Database of the Global Environment. Utrecht University https://doi.org/10.24416/UU01-94FNH0 (2024).

24. Tang, F. H. M. *et al.* CROPGRIDS: A global geo-referenced dataset of 173 crops. *Scientific Data* **11**, 413 (2024).
25. Parente, L. *et al.* Annual 30-m maps of global grassland class and extent (2000-2022) based on spatiotemporal Machine Learning. *Sci. Data* **11**, 1303 (2024).
26. Reinermann, S., Asam, S. & Kuenzer, C. Remote sensing of grassland production and management—A review. *Remote Sens. (Basel)* **12**, 1949 (2020).
27. Ball, P. R. & Ryden, J. C. Nitrogen relationships in intensively managed temperate grasslands. *Plant Soil* **76**, 23–33 (1984).
28. Council of the European Communities. Council Directive 91/676/EEC of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources. *Official Journal of the European Communities* **L 375**, 1–8 (1991).
29. Friedl, M. & Sulla-Menashe, D. MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006. NASA Land Processes Distributed Active Archive Center https://doi.org/10.5067/MODIS/MCD12C1.006 (2015).
30. Sulla-Menashe, D. & Friedl, M. A. User Guide to Collection 6 MODIS Land Cover (MCD12Q1 and MCD12C1) Product. Preprint at https://landweb.modaps.eosdis.nasa.gov/data/userguide/MCD12Q1Userguide04042018.pdf (2018).
31. Zhang, B. *et al.* Global manure nitrogen production and application in cropland during 1860–2014: a 5 arcmin gridded global dataset for Earth system modeling. *Earth Syst. Sci. Data* **9**, 667–678 (2017).
32. Maffenini, L. *et al.* GHS-POPWARP User Guide. https://publications.jrc.ec.europa.eu/repository/handle/JRC132758 (2023).
33. Yang, J., Tian, H. & Büchner, M. ISIMIP3a Nitrogen Deposition Input Data. ISIMIP Repository https://doi.org/10.48364/ISIMIP.759077.4 (2026).

34. Zhang, X. *et al.* Quantification of global and national nitrogen budgets for crop production. *Nat. Food* **2**, 529–540 (2021).
35. Ludemann, C. I. *et al.* A global FAOSTAT reference database of cropland nutrient budgets and nutrient use efficiency (1961–2020): nitrogen, phosphorus and potassium. *Earth Syst. Sci. Data* **16**, 525–541 (2024).
36. Calvo Buendia, E. *et al. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. https://beta.ghg.inergy.vn/upload/others/2025/03/27/19R_V5_IPCC%202019_Waste_B%E1%BA%A3n%20%C4%91%E1%BA%A7y%20%C4%91%E1%BB%A7-94ef79b7-a22f-4339-8c0c-36763a149144.pdf (2019).
37. Nielsen, O.-K. EMEP/EEA air pollutant emission inventory guidebook 2013. Technical guidance to prepare national emission inventories. https://www.osti.gov/etdeweb/biblio/22128042 (2013).
38. Bouwman, A. F., Boumans, L. J. M. & Batjes, N. H. Estimation of global NH_3 volatilization loss from synthetic fertilizers and animal manure applied to arable lands and grasslands: AMMONIA EMISSION FROM FERTILIZERS. *Global Biogeochem. Cycles* **16**, 8–1–8–14 (2002).
39. Ramankutty, N., Evan, A. T., Monfreda, C. & Foley, J. A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochem. Cycles* **22**, (2008).
40. Nguyen, T. H. *et al.* NPKGRIDS: a global georeferenced dataset of N, P2O5, and K2O fertilizer application rates for 173 crops. *Sci. Data* **11**, 1179 (2024).
41. Gavrilova, O. *et al.* Emissions from livestock and manure management. in *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories* (2019).
42. Dong, H. *et al. Emissions from Livestock and Manure Management*. (IPCC, 2006).
43. De Klein, C. *et al.* N2O emissions from managed soils, and CO2 emissions from lime and urea application. in *2006 IPCC guidelines for national greenhouse gas inventories* (2006).

44. Ludemann, C. I. *et al.* Global data on crop nutrient concentration and harvest indices. Dryad https://doi.org/10.5061/DRYAD.N2Z34TN0X (2023).
45. Hong, B., Swaney, D. P. & Howarth, R. W. Estimating net anthropogenic nitrogen inputs to U.S. watersheds: Comparison of methodologies. *Environmental Science and Technology* **47**, 5199–5207 (2013).
46. Han, H. & Allan, J. D. Estimation of nitrogen inputs to catchments: comparison of methods and consequences for riverine export prediction. *Biogeochemistry* **91**, 177–199 (2008).
47. Byrnes, D. K., Van Meter, K. J. & Basu, N. B. Trajectories Nutrient Dataset for Nitrogen (TREND-Nitrogen) v2. https://doi.org/10.6084/m9.figshare.20915989.v1 (2022).
48. Peoples, M. B., Giller, K. E., Jensen, E. S. & Herridge, D. F. Quantifying country-to-global scale nitrogen fixation for grain legumes: I. Reliance on nitrogen fixation of soybean, groundnut and pulses. *Plant Soil* **469**, 1–14 (2021).
49. FAO. FAOSTAT Statistical Database. Food and Agriculture Organization of the United Nations (2024).
50. Herridge, D. F. Biological nitrogen fixation: Techniques to enhance. *Encycl Soil Sci* https://api.taylorfrancis.com/content/chapters/edit/download?identifierName=doi&identifierValue=10.1201/NOE0849338304-120042636&type=chapterpdf (2006).
51. Herridge, D. F., Giller, K. E., Jensen, E. S. & Peoples, M. B. Quantifying country-to-global scale nitrogen fixation for grain legumes II. Coefficients, templates and estimates for soybean, groundnut and pulses. *Plant Soil* **474**, 1–15 (2022).
52. Bouwman, L. *et al.* Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900-2050 period. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20882–20887 (2013).
53. Beusen, A. H. W., Van Beek, L. P. H., Bouwman, A. F., Mogollón, J. M. & Middelburg, J. J. Coupling

global models for hydrology and nutrient loading to simulate nitrogen and phosphorus retention in surface water – description of IMAGE–GNM and analysis of performance. *Geosci. Model Dev.* **8**, 4045–4067 (2015).

54. FAO. Global livestock environmental assessment model, Version 2.0. Preprint at https://www.fao.org/fileadmin/user_upload/gleam/docs/GLEAM_2.0_Model_description.pdf (2018).
55. Herrero, M. *et al.* Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 20888–20893 (2013).
56. Einarsson, R. *et al.* Crop production and nitrogen use in European cropland and grassland 1961-2019. *Sci. Data* **8**, 288 (2021).
57. Whitehead, D. C. *Nutrient Elements in Grassland*. (CABI Publishing, 2000).
58. Soussana, J.-F. & Lemaire, G. Coupling carbon and nitrogen cycles for environmentally sustainable intensification of grasslands and crop-livestock systems. *Agric. Ecosyst. Environ.* **190**, 9–17 (2014).
59. Swaney, D. P., Howarth, R. W. & Hong, B. County, subregional and regional nitrogen data derived from the Net Anthropogenic Nitrogen Inputs (NANI) toolbox. *Data Brief* **18**, 1877–1888 (2018).
60. Swaney, D. P., Howarth, R. W. & Hong, B. Nitrogen use efficiency and crop production: Patterns of regional variation in the United States, 1987-2012. *Sci. Total Environ.* **635**, 498–511 (2018).
61. Fowler, D. *et al.* The global nitrogen cycle in the twenty-first century. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20130164 (2013).
62. Liu, X. *et al.* Enhanced nitrogen deposition over China. *Nature* **494**, 459–462 (2013).
63. Eurostat. Gross nutrient balance. *Gross nutrient balance* Eurostat https://doi.org/10.2908/AEI_PR_GNB (2023).
64. Mariotti, F., Tomé, D. & Mirand, P. P. Converting nitrogen into protein--beyond 6.25 and Jones' factors. *Crit. Rev. Food Sci. Nutr.* **48**, 177–184 (2008).

65. Van Puijenbroek, P., Beusen, A. H. & Bouwman, A. F. Global nitrogen and phosphorus in urban waste water based on the Shared Socio-economic pathways. *Journal of environmental management* **231**, 446–456 (2019).
66. Bolt, J. & van Zanden, J. L. Maddison-style estimates of the evolution of the world economy: A new 2023 update. *J. Econ. Surv.* **39**, 631–671 (2025).
67. United Nations Statistics Division. *Population Connected to Wastewater Treatment*. https://unstats.un.org/unsd/environment/qindicators.htm (2024).
68. Wastewater - connection rates to treatment. https://data-explorer.oecd.org/vis?df[ds]=DisseminateFinalDMZ&df[id]=DSD_WATER_TREAT%40DF_WATER_TREAT&df[ag]=OECD.ENV.EPI&dq=AUS.A..&pd=2010,&to[TIME_PERIOD]=false.
69. United Nations Statistics Division. *Population Connected to Wastewater Collecting System*. https://unstats.un.org/unsd/environment/qindicators.htm (2024).
70. Metcalf and Eddy Inc. *Wastewater Engineering*. (McGraw-Hill Education (ISE Editions), London, England, 1991).
71. Trimmer, J. T. & Guest, J. S. Recirculation of human-derived nutrients from cities to agriculture across six continents. *Nat. Sustain.* **1**, 427–435 (2018).
72. Dai, Z. Developing energy and nutrient mass balances to inform value recovery options in municipal wastewater treatment systems. (Newcastle University, 2019).
73. Sagberg, P., Ryrfors, P. & Berg, K. G. The mass balances of nitrogen and carbon in a compact nitrogen and phosphorus WWTP. in *Chemical Water and Wastewater Treatment V* 231–242 (Springer Berlin Heidelberg, Berlin, Heidelberg, 1998). doi:10.1007/978-3-642-72279-0_19.
74. Shi, C. Y. *Mass Flow and Energy Efficiency of Municipal Wastewater Treatment Plants*. (IWA Publishing, London, England, 2011).
75. EurEau, A. Europe's water in figures—An overview of the European drinking water and waste water

sectors. *The European Federation of National Water Associations Brussels* (2017).

76. Collivignarelli, M. C., Bellazzi, S., Calabria, L. M. R., Torretta, V. & Rada, E. C. Wastewater reuse for agriculture in Europe: regulatory frameworks, treatment technologies, and socio-economic perspectives. *Curr. Opin. Environ. Sustain.* **80**, 101630 (2026).
77. Beecher, N. *et al. National Biosolids Data Project: Biosolids management in the U. S.* (2022).
78. Donatello, S. & Cheeseman, C. R. Recycling and recovery routes for incinerated sewage sludge ash (ISSA): a review. *Waste Manag.* **33**, 2328–2340 (2013).
79. Werther, J. & Ogada, T. Sewage sludge combustion. *Prog. Energy Combust. Sci.* **25**, 55–116 (1999).
80. Noda, K. Night soil to chemical fertilizer: Lessons from the abolishment of the circular economy in postwar japan. *Circ. Econ. Sustain.* https://doi.org/10.1007/s43615-024-00354-w (2024) doi:10.1007/s43615-024-00354-w.
81. Noda, K. Farmland to sea: the contemporary policy history of night soil management in Tokyo. *Water Hist.* **16**, 291–306 (2024).
82. King, F. H. *Farmers of Forty Centuries: Organic Farming in China, Korea, and Japan*. (Dover Publications, New York, NY, 2004).
83. Shirai, Y., Leisz, S. J. & Kyuma, K. A short history of the utilization of nightsoil in agriculture. *Sanitation* **7**, 1–24 (2023).
84. Ferguson, D. T. Nightsoil and the 'Great Divergence': human waste, the urban economy, and economic productivity, 1500–1900. *J. Glob. Hist.* **9**, 379–402 (2014).
85. Aliahmad, A. *Rethinking Wastewater: Sustainability Transition Assessment of Nutrient Recycling Systems from Source-Separated Wastewater : Focused on Urine Concentration and Recycling*. (Swedish University of Agricultural Sciences, 2025). doi:10.54612/a.ms2ndhdmba.
86. Miörner, J. & Binz, C. Towards a multi-scalar perspective on transition trajectories. *Environ. Innov. Soc. Transit.* **40**, 172–188 (2021).

87. Starck, T., Fardet, T. & Esculier, F. Fate of nitrogen in French human excreta: Current waste and agronomic opportunities for the future. *Sci. Total Environ.* **912**, 168978 (2024).
88. Yang, X. M., Morita, A., Nakano, I., Kushida, Y. & Ogawa, H. History and current situation of night soil treatment systems and decentralized wastewater treatment systems in Japan. *Water Pract. Technol.* **5**, wpt2010096–wpt2010096 (2010).
89. Gunjal, K., Goodbody, S., Coslet, C. & Hollema, S. FAO/WFP crop and food security assessment mission to the Democratic People's Republic of Korea, November 2012. https://openknowledge.fao.org/items/52b13042-b91e-4dd4-82a2-d7972c293e75 (2012).
90. Son Hyemin for RFA Korean. North Koreans fight over feces as annual 'compost battle' begins. *Radio Free Asia* https://www.rfa.org/english/korea/2025/01/09/north-korea-annual-manure-battle/ (2025).
91. Lee, M. C. Yearly 'battle' begins in North Korea over human waste for fertilizer. *Radio Free Asia* https://www.rfa.org/english/news/korea/manure-02012022130916.html (2022).
92. Citizens Fight Over Feces to Fill Human Fertilizer Quota in North Korea. *Radio Free Asia* https://www.rfa.org/english/news/korea/nk-manure-quota-2020-10142019154836.html (2019).
93. Pham, H. G., Harada, H., Fujii, S., Nguyen, P. H. L. & Huynh, T. H. Transition of human and livestock waste management in rural Hanoi: a material flow analysis of nitrogen and phosphorus during 1980–2010. *J. Mater. Cycles Waste Manag.* **19**, 827–839 (2017).
94. Tran-Thi, N. *et al.* Turning poop into profit: Cost-effectiveness and soil transmitted helminth infection risk associated with human excreta reuse in Vietnam. *PLoS Negl. Trop. Dis.* **11**, e0006088 (2017).
95. Jensen, P. K. M., Phuc, P. D., Knudsen, L. G., Dalsgaard, A. & Konradsen, F. Hygiene versus fertiliser: the use of human excreta in agriculture--a Vietnamese example. *Int. J. Hyg. Environ. Health* **211**, 432–439 (2008).

96. Chan-Yeung, M. The business of night soil collection. https://www.hkmj.org/abstracts/v22n5/518.htm (2016).
97. Rajbhandari, K. Ecological Sanitation Latrines. https://nl.ircwash.org/sites/default/files/Rajbhandari-2008-Ecological.pdf (2008).
98. Byrnes, D. K., Chang, S. Y., Basu, N. B. & Van Meter, K. TREND-Phosphorus and gTREND-Phosphorus - Long-term phosphorus mass balance data for the contiguous United States (1930-2017). Figshare+ https://doi.org/10.25452/FIGSHARE.PLUS.30158131.V1 (2026).
99. Van Meter, K. J. *et al.* Beyond the mass balance: Watershed phosphorus legacies and the evolution of the current water quality policy challenge. *Water Resour. Res.* **57**, (2021).
100. Dijkstra, L. *et al.* Applying the Degree of Urbanisation to the globe: A new harmonised definition reveals a different picture of global urbanisation. *J. Urban Econ.* **125**, 103312 (2021).
101. Muleta, M. K. & Nicklow, J. W. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *J. Hydrol.* **306**, 127–145 (2005).
102. Mishra, S. Uncertainty and sensitivity analysis techniques for hydrologic modeling. *J. Hydroinformatics* **11**, 282–296 (2009).
103. FAO. Agriculture organization of the united nations: Food energy—methods of analysis and conversion factors. *FAO Food Nutr. Pap.* (2003).
104. Forbes, H. & Others. Food waste index report 2021. https://www.sidalc.net/search/Record/oai:wedocs.unep.org:20.500.11822-35280/Description (2021).
105. Frey, C., Penman, J., Hanle, L., Monni, S. & Ogle, S. *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. (2006).
106. Solazzo, E. *et al.* Uncertainties in the Emissions Database for Global Atmospheric Research (EDGAR) emission inventory of greenhouse gases. *Atmos. Chem. Phys.* **21**, 5655–5683 (2021).

107. Gotway, C. A. & Young, L. J. Combining incompatible spatial data. *J. Am. Stat. Assoc.* **97**, 632–648 (2002).
108. Brehob, M. M. *et al.* The US EPA's national nutrient inventory: Critical shifts in US nutrient pollution sources from 1987 to 2017. *Environ. Sci. Technol.* **59**, 27836–27852 (2025).
109. Li, J. *et al.* Uncovering the spatial characteristics of global net anthropogenic nitrogen input at high resolution and across 1.42 million lake basins. *Sci. Total Environ.* **953**, 176143 (2024).
110. Tuholske, C. *et al.* Mapping global inputs and impacts from of human sewage in coastal ecosystems. *PLoS One* **16**, e0258898 (2021).
111. Sabo, R. Next generation National Nutrient Inventory for nitrogen and phosphorus (1987 to 2017). U.S. EPA Office of Research and Development (ORD) https://doi.org/10.23719/1532436 (2025).
112. NADP. National Atmospheric Deposition Program. *National Trends Network* https://nadp.slh.wisc.edu/networks/national-trends-network/.