Table S1. Global mean soil N_2O emissions (g N m^{-2} yr^{-1}) for the 1850s and 2010s from eight NMIP2 terrestrial biosphere models under the SH1 experiment, and the change (Δ = 2010s – 1850s).

Model	Global mean soil N_2O emissions		Δ	Grid cells with Δ > 0 (%)
	1850s	2010s		
CLASSIC	0.026	0.069	+0.043	89.6%
DLEM	0.042	0.067	+0.026	66.2%
ELM	0.057	0.083	+0.026	79.5%
ISAM	0.029	0.058	+0.029	76.0%
LPX-Bern	0.038	0.063	+0.025	79.1%
O-CN	0.034	0.053	+0.019	61.4%
ORCHIDEE	0.057	0.072	+0.015	59.9%
VISIT	0.050	0.074	+0.023	77.0%
Ensemble mean	0.042	0.067	+0.026	-

The rightmost column shows the percentage of valid grid cells where Δ>0. All values are spatial means across valid grid cells for each model.

Table S2. Sensitivity of robustness classification to agreement threshold (κ_A) and land-use fraction thresholds.

Land use	Threshold	n	κ_A>=0.75 (6/8)				κ_A>=0.875 (7/8)			
			Robust-increase (%)	Robust-decrease (%)	Divergent (%)	Uncertain (%)	Robust-increase (%)	Robust-decrease (%)	Divergent (%)	Uncertain (%)
Global	-	56,852	40.8	2.6	36.1	**20.5**	26.0	0.5	57.1	**16.4**
Cropland	>=0.2	11,251	27.0	0.0	5.2	**67.8**	24.2	0.0	15.4	**60.4**
Cropland	>=0.3[a]	8,203	22.7	0.0	4.6	**72.6**	20.5	0.0	14.0	**65.5**
Cropland	>=0.5	4,099	17.6	0.0	4.1	**78.2**	15.4	0.0	13.9	**70.7**
Forest	>=0.3	20,100	33.1	2.9	**46.8**	17.2	16.7	0.3	**70.8**	12.2
Forest	>=0.5[a]	18,175	33.8	3.1	**50.0**	13.1	16.3	0.4	**74.8**	8.6
Pasture	>=0.3	16,003	**59.1**	1.2	17.4	22.3	**43.8**	0.2	39.4	16.6

[a]Main analysis settings: $\kappa_A = 0.75$ ($\geq$ 6 or 8 models agree on sign), Cropland $\geq$ 0.3, Forest $\geq$ 0.5, Pasture $\geq$ 0.3.

SD_{Q75} = 0.0392 (75th percentile of inter-model standard deviation across all 56,852 valid grid cells; held fixed for all threshold combinations). Pasture threshold is not varied because the main findings are insensitive to this parameter.

Classification: Robust-increase ($\kappa_A \geq$ threshold, SD < SD_{Q75}, $\overline{\Delta} > 0$); Robust-decrease (same but $\overline{\Delta} < 0$); Divergent (κ_A < threshold); Uncertain ($\kappa_A \geq$ threshold, SD $\geq SD_{Q75}$).

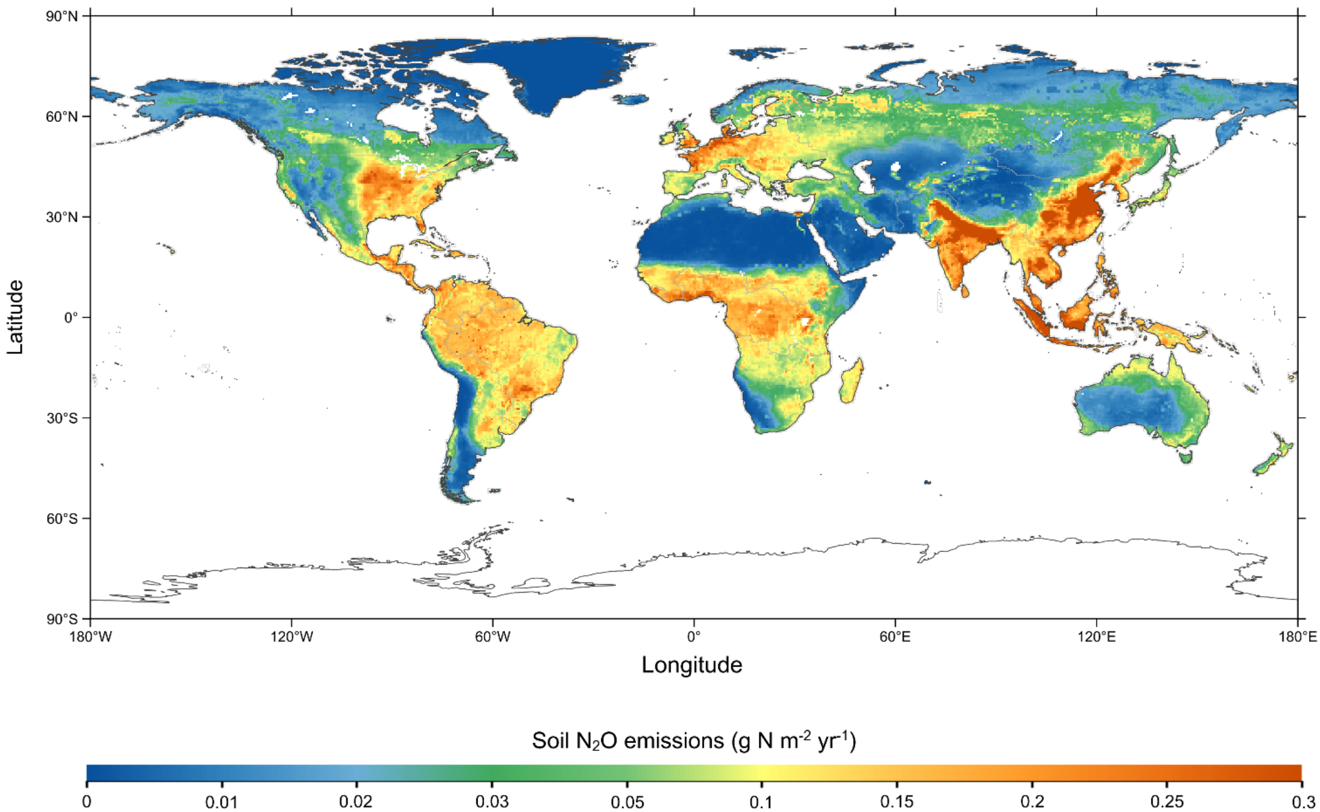

Figure S1. Ensemble-mean soil N_2O emissions for the 2010s (g N m^{-2} yr^{-1}) from eight NMIP2 terrestrial biosphere models under the SH1 experiment. The color scale is logarithmic to capture the wide range of emission rates across regions. High-emission areas correspond to intensively managed agricultural regions (e.g. eastern China, the Indo-Gangetic Plain, western Europe, and the US Midwest), while lower emissions are found in boreal and arid regions.

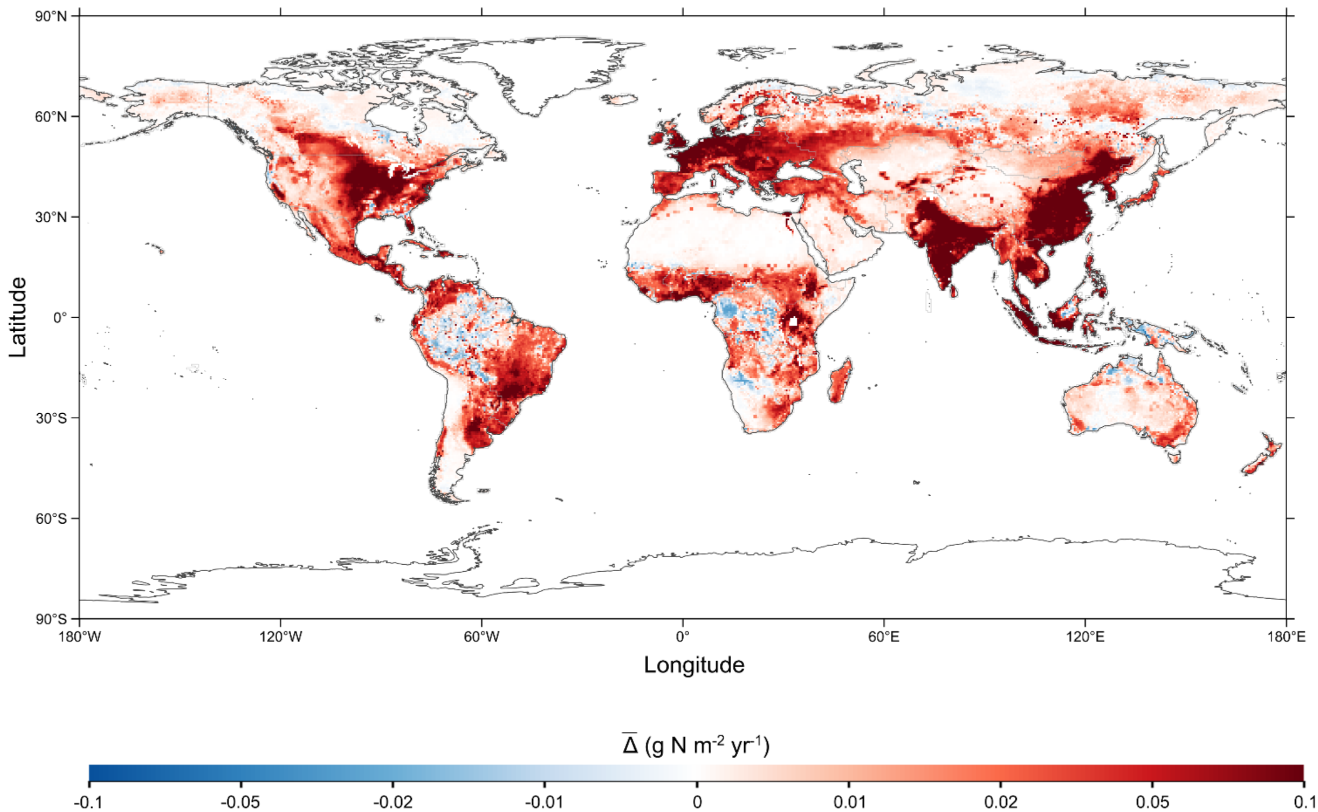

Figure S2. Spatial distribution of ensemble-mean change in soil N_2O emissions ($\overline{\Delta}$; g N m^{-2} yr^{-1}) between the 1850s and 2010s from eight NMIP2 models. Red shading indicates increased emissions and blue shading indicates decreased emissions. The color scale is logarithmic. Most land areas show positive $\overline{\Delta}$, with the largest increases concentrated in regions of intensive agricultural activity.

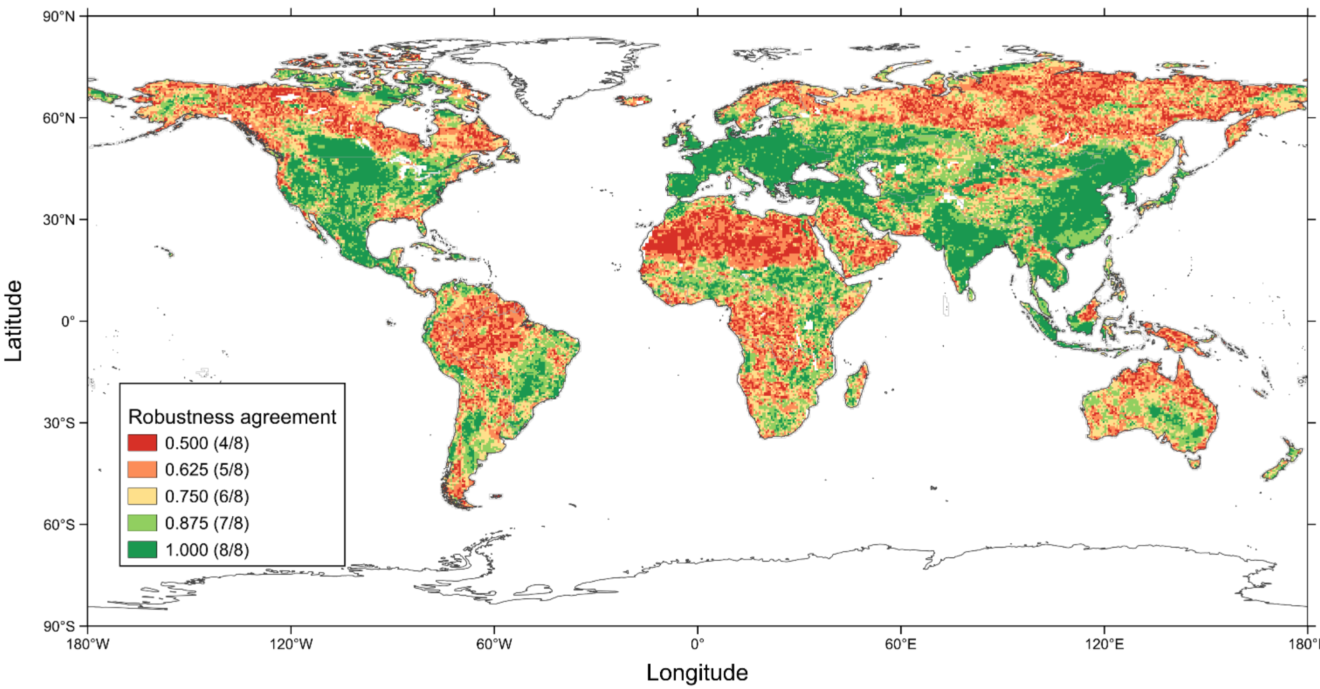

Figure S3. Spatial distribution of model agreement (κ_A) on the sign of N_2O emission change between the 1850s and 2010s. Agreement is defined as the fraction of models that agree on the dominant sign of change, ranging from 0.500 (4/8, no consensus) to 1.000 (8/8, full consensus). High agreement (green) is prevalent in agricultural regions and boreal zones, while low agreement (red/orange) is widespread in tropical and subtropical forests.

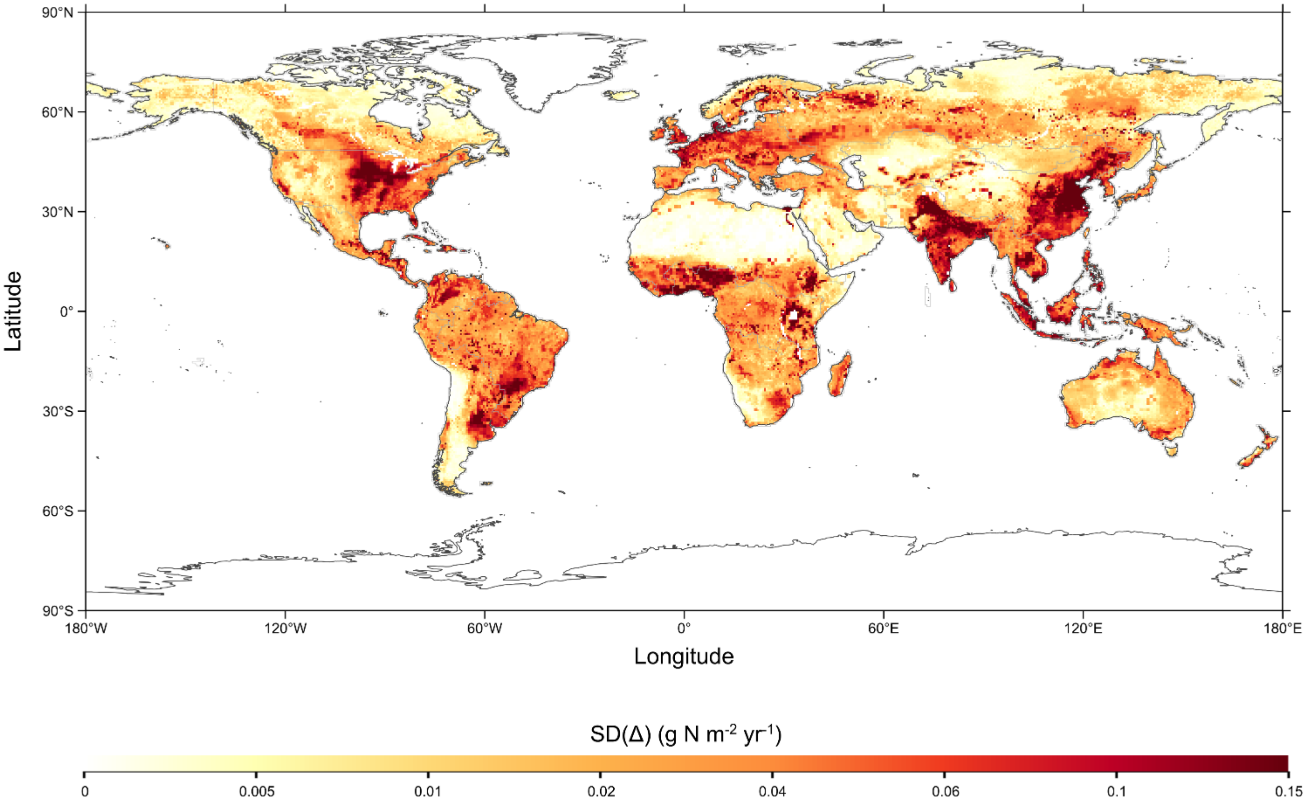

Figure S4. Spatial distribution of inter-model standard deviation of changes (SD (Δ); g N m^{-2} yr^{-1}) in soil N_2O emissions between the 1850s and 2010s across eight NMIP2 models. The color scale is logarithmic. Large inter-model spread is concentrated in regions with high anthropogenic nitrogen inputs, such as eastern China, South and Southeast Asia, and parts of western Europe and North America.

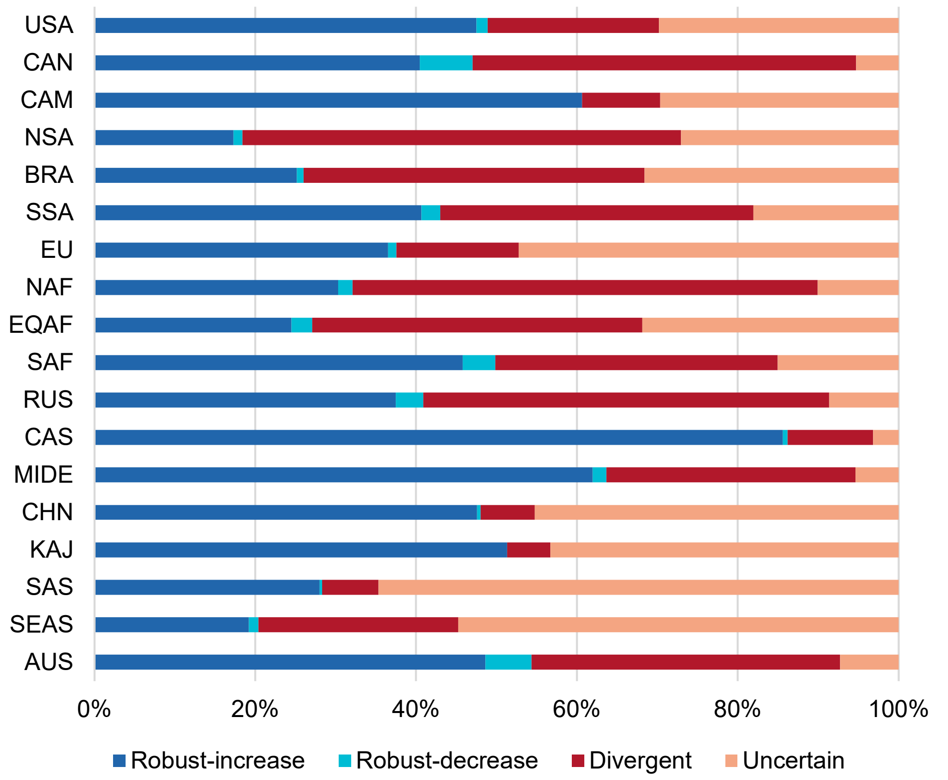

Figure S5. Robustness classification composition by region. The 18 land regions follow the classification used in Global Carbon Project greenhouse gas budget assessments (Saunois et al., 2020; Tian et al., 2020, 2024), constructed using geopolitical boundaries on the basis of size, geopolitical importance, and vegetation type. Bars show the percentage of valid grid cells in each region classified as Robust-increase (dark blue), Robust-decrease (cyan), Divergent (red), and Uncertain (orange). Region abbreviations: USA, United States; CAN, Canada; CAM, Central America; NSA, Northern South America; BRA, Brazil; SSA, Southwestern South America; EU, Europe; NAF, Northern Africa; EQAF, Equatorial Africa; SAF, Southern Africa; RUS, Russia; CAS, Central Asia; MIDE, Middle East; CHN, China; KAJ, Korea and Japan; SAS, South Asia; SEAS, Southeast Asia; AUS, Australasia.