Supplement of Biogeosciences, 22, 7845–7863, 2025
https://doi.org/10.5194/bg-22-7845-2025-supplement

Biogeosciences Open Access EGU

Supplement of

Using explainable AI to diagnose the representation of environmental drivers in process-based soil organic carbon models

Lingfei Wang et al.

Correspondence to: Lingfei Wang (lingfei.wang@unsw.edu.au)

S1. Vertically resolved MIMICS

In this study, we used the vertically resolved version of MIMICS developed by Wang et al. (2020). Since soil moisture is not represented in the original version of the model, we introduced a soil moisture scalar $K(\theta, \psi)$ to the model,

$$K(\theta, \psi) = e^{\lambda\psi} \times \left(k_{a.min} + (1 - k_{a.min}) \times \left(\frac{\phi-\theta}{\phi}\right)^{0.5}\right) \times \left(\frac{\theta}{\phi}\right)^{0.5} \quad \text{(S1)}$$

where $e^{\lambda\psi}$ describes the dependence of soil microbial activities on water potential, and λ (kPa^{-1}) is the dependence of biological activity on soil matric potential ψ (kPa). $\left(k_{a.min} + (1 - k_{a.min}) \times \left(\frac{\phi-\theta}{\phi}\right)^{0.5}\right)$ represents the effect of oxygen limitation on microbial activities, and the parameter $k_{a.min}$ is the minimum relative SOC decomposition rate when the soil is fully saturated, and the oxygen supply rate is at its lowest. $\left(\frac{\theta}{\phi}\right)^{0.5}$ represents the effect of diffusion limitation of substrates on SOC decomposition, and θ is soil moisture content (mm mm^{-3}) and ϕ is soil porosity (mm mm^{-3}). More details can be found at (Ghezzehei et al., 2019). We multiplied both V_{max} and microbial turnover rate (τ) by $K(\theta, \psi)$ to represent the effects of soil moisture on microbial growth rates, which include both uptake and turnover (Wieder et al., 2019).

In addition, a depth-dependent decline in microbial activity is introduced following an exponential decay function (Koven et al., 2013),

$$r_z = \exp(-z \times z_\tau) \quad \text{(S2)}$$

Where r is depth-dependent scalar, z is soil depth (cm), and z_τ (default value 2.0) is the e-folding depth of intrinsic turnover rates. This scalar was applied to the microbial maximum reaction velocity (V_{max}, see E1 below) to reduce SOC turnover rates with increasing depth.

S2. Structure and equations of MES-C

MES-C has two litter pools, metabolic litter (LIT_m) and structural litter (LIT_s), and the decomposition of litter and SOC are governed by two functional types of microbes including *r*-selected (MIC_r) and *K*-selected (MIC_k) microbes. SOC is divided into three parts: light molecule weight carbon (LMWC), mineral associated organic carbon (MAOC) and aggregated organic carbon, where aggregated carbon includes three fractions: aggregated metabolic carbon (AGG_m), aggregated structural carbon (AGG_s) and aggregated mineral associated carbon (AGG_{maoc}). Further description and assumptions can be found below or at Wieder et al., 2015 and Abramoff et al., 2022.

S2.1 Metabolic and structural litter pool – LIT_m and LIT_s

The representation of litter pools in MES-C is based on litter chemistry and decomposition dynamic, and fresh litter is partitioned into high- and low-quality litter pools (LIT_m and LIT_s, respectively) following a linear function of litter nitrogen to lignin ratios (f_{met}; Table S1). In this study, we parameterized this linear function using assumed litter quality values. Specifically, we set lignin/C ratios to 0.15 for grasses and 0.25 for woody plants, respectively. We set C/N ratios to 35 for aboveground litter and 60 for belowground litter, respectively. And $lignin/N$ ratios were then calculated as the product of lignin/C and C/N. These values differ between vegetation types (grasses vs. woody) and litter origin (above- vs. belowground) to reflect in litter quality. A fraction of litter inputs (f_m and fs for metabolic and structural litter, respectively; Table S1) bypasses both litter pools and microbial biomass pools, and it's directly aggregated into corresponding SOC pools (e.g., AGG_m and AGG_s).

It's assumed that ecosystem carbon cycle is at steady state, and litter input is represented by above- and below-ground net primary production (NPP) in this study. As we simulate SOC turnover at multiple soil depths, the vertical allocation of below-ground NPP in soils is determined by a negative exponential function describing the decline of root distribution with depth (Jackson et al., 1996),

$$Y = 1 - \beta^d \quad \text{(S3)}$$

Where Y is the cumulative root fraction from the soil surface to depth d (cm), and β is the fitted parameter. In this study we applied the same β (0.966) for all vegetation types. This approach may introduce uncertainties in SOC predictions at depth. However, due to the lack of consistent parameters describing root distributions for different vegetation types, and because approximately 70% of roots are generally concentrated in the top 30 cm of soil across most vegetation types (Jackson et al., 1996), we expect the resulting uncertainties to be limited, particularly in the top 30 cm, which is the main focus of our study.

S2.2 Microbial biomass pools – MIC_r and MIC_k

In MES-C, microbial biomass pools control the turnover of both litter and SOC, representing microbial functional types that follow distinct growth strategies—copiotrophic (r-selected) and oligotrophic (K-selected), as in the MIMICS framework (Wieder et al., 2015). These microbial communities are characterized by key functional traits that define their growth strategies, including microbial kinetics (described by reversed Michaelis–Menten kinetics), microbial growth efficiency (MGE), and microbial turnover rate (τ) (Table S1).

S2.3 Low-molecular-weight carbon pool – LMWC

In MES-C, LMWC represents the low molecular mass, soluble organic compounds which are readily accessible to microbial uptake. LMWC receives inputs via the decomposition of the litter pools through microbially mediated breakdown, and also serves as the immediate substrate for microbial biomass. The turnover of LMWC is rapid relative to more protected pools such as MAOC and aggregated carbon, and it is depleted by microbial assimilation and leaching losses.

S2.4 Mineral-associated organic carbon pool – MAOC

The MAOC pool represents the fraction of SOC that is bound to mineral surfaces such as clay particles and metal oxides. This pool is generally characterized by slow turnover rates, with residence times ranging from decades to millennia (Six et al., 2002). In the MES-C model, MAOC originates from two main sources. The first is the adsorption of microbial necromass following microbial death. The second is the direct sorption of low-molecular-weight carbon (LMWC) compounds, which follows a Langmuir-type saturating relationship. The maximum sorption capacity is determined by the soil's clay and silt content and is further modulated by soil pH (Abramoff et al., 2022). Additionally, MAOC can become physically protected within soil aggregates and later released upon aggregate breakdown, which are represented in the model by the rate constants k_{ma} and kab for aggregate formation and breakdown, respectively (Table S1).

S2.5 Aggregated organic carbon pool – AGG_m, AGG_s and AGG_{maoc}

In MES-C, the aggregates are not represented as an independent pool, consistent with the framework of Laub et al. (2024). Instead, aggregates are composed of existing carbon pools, including LIT_m, LIT_s, and MAOC, which become physically protected from decomposition upon incorporation into the aggregate matrix. Upon aggregate break down, each carbon fraction is returned to its original pool without any loss of mass.

S2.6 Vertical transfer of SOC in MES-C

The vertical transport of SOC via bioturbation and diffusion in same as that in vertically resolved MIMICS as described above.

S2.7 Equations of MES-C

In MES-C, decomposition of litter follows forward Michaelis-Menten kinetics, and microbial uptake of LMWC follows reversed Michaelis-Menten kinetics. Both kinetics are dependent on temperature sensitive maximum reaction velocities and half saturation constants calculated as,

$$V_{max} = e^{(V_{slope} \times T + V_{int})} \times a_v \times V_{mod} \quad \text{(S4)}$$
$$K_m = e^{(K_{slope} \times T + K_{int})} \times a_k \times K_{mod} \quad \text{(S5)}$$

The fluxes from donor to receiver pools are calculated as,

$$LIT_m \rightarrow MIC_r = LIT_m \times V_{max[r1]} \times K(\theta, \psi) \times MIC_r / (K_{m[r1]} + MIC_r) \quad \text{(S6)}$$
$$LIT_s \rightarrow MIC_r = LIT_s \times V_{max[r2]} \times K(\theta, \psi) \times MIC_r / (K_{m[r2]} + MIC_r) \quad \text{(S7)}$$
$$LMWC \rightarrow MIC_r = MIC_r \times V_{max[r3]} \times K(\theta, \psi) \times LMWC / (K_{m[r3]} + LMWC) \quad \text{(S8)}$$
$$MIC_r \rightarrow SOC = \mu_{mic} \times MIC_r \times {\tau_r}^2 \quad \text{(S9)}$$
$$LIT_m \rightarrow MIC_k = LIT_m \times V_{max[k1]} \times K(\theta, \psi) \times MIC_k / (K_{m[k1]} + MIC_k) \quad \text{(S10)}$$
$$LIT_s \rightarrow MIC_k = LIT_s \times V_{max[k2]} \times K(\theta, \psi) \times MIC_k / (K_{m[k2]} + MIC_k) \quad \text{(S11)}$$
$$LMWC \rightarrow MIC_k = MIC_k \times V_{max[k3]} \times K(\theta, \psi) \times LMWC / (K_{m[k3]} + LMWC) \quad \text{(S12)}$$
$$MIC_k \rightarrow SOC = \mu_{mic} \times MIC_k \times {\tau_k}^2 \quad \text{(S13)}$$
$$LMWC \rightarrow MAOC = K(\theta, \phi) \times k_{adsorp} \times LMWC \times (1 - MAOC / Q_{max}) \quad \text{(S14)}$$
$$MAOC \rightarrow LMWC = k_{desorp} \times MAOC / Q_{max} \quad \text{(S15)}$$
$$AGG_m \rightarrow LIT_m = k_{ab} \times K(\theta, \phi) \times AGG_m \quad \text{(S16)}$$
$$AGG_s \rightarrow LIT_s = k_{ab} \times K(\theta, \phi) \times AGG_s \quad \text{(S17)}$$
$$AGG_{maoc} \rightarrow MAOC = k_{ab} \times K(\theta, \phi) \times AGG_{maoc} \quad \text{(S18)}$$
$$MAOC \rightarrow AGG_{maoc} = k_{ma} \times K(\theta, \phi) \times MAOC \quad \text{(S19)}$$
$$LMWC \rightarrow F_l = k_l \times K(\theta, \phi) \times LWMC \quad \text{(S20)}$$

Thus, changes in C pools can be described using the following equations,

$$\frac{dLIT_m}{dt} = I_m \times (1 - f_m) - S6 - S10 + S16 \quad \text{(S21)}$$
$$\frac{dLIT_s}{dt} = I_s \times (1 - f_s) - S7 - S11 + S17 \quad \text{(S22)}$$
$$\frac{dMIC_r}{dt} = (MGE[1] \times S6) + (MGE[2] \times S7) + (MGE[1] \times S8) - S9 \quad \text{(S23)}$$
$$\frac{dMIC_k}{dt} = (MGE[3] \times S10) + (MGE[4] \times S11) + (MGE[3] \times S12) - S13 \quad \text{(S24)}$$
$$\frac{dLMWC}{dt} = f_a[r] \times S11 + f_a[k] \times S13 + S15 - S8 - S12 - S14 - S20 \quad \text{(S25)}$$
$$\frac{dMAOC}{dt} = f_c[r] \times S9 + f_c[k] \times S13 + S18 + S14 - S15 - S19 \quad \text{(S26)}$$
$$\frac{dAGG_m}{dt} = I_m \times f_m - S16 \quad \text{(S27)}$$
$$\frac{dAGG_s}{dt} = I_s \times f_s - S17 \quad \text{(S28)}$$
$$\frac{dAGG_{maoc}}{dt} = S19 - S18 \quad \text{(S29)}$$

S3. Technical implementation of vertically resolved MIMICS and MES-C

Both the vertically resolved MIMICS and MES-C models are implemented in Fortran and operate at an hourly timestep. The systems of differential equations are numerically integrated using a fourth-order Runge–Kutta (RK4) solver. SOC transfer in the vertical direction is simulated by solving the diffusion equation (Eqn1 in Wang et al., 2021) numerically using Crank-Nicholson method. Simulations of SOC dynamics are performed in 10 cm depth increments down to 1.5 m, and model outputs are averaged every 30 cm for comparison with observations. We assume that SOC at each site is at equilibrium, and both models are spun up for 1,000 years, with equilibrium assumed when the relative change in carbon pool size between two successive spin-up iterations is less than 0.05%.

Parameter optimization was conducted using the global optimization algorithm shuffled complex evolution (SCE-UA, version 2.2) (Duan et al., 1993) by minimising the residual sum of squares between the observed and modelled values at all depths. For sites lacking SOC observations in deeper layers (Figure 1), the model was still run for the full soil profile but these deeper layers were excluded from the cost function during parameter optimization to ensure that calibration was based only on available observations. The optimization was executed on a GPU platform to enhance computational efficiency. For MIMICS, we optimized parameters previously identified as influential to model performance (Wang et al., 2020; Zhang et al., 2020). For MES-C, we first conducted a Morris sensitivity analysis following Lu et al. (2013) for 12 clusters (see Section 2.3 in main text), and selected 6 relatively sensitive parameters across all clusters. The results of parameter sensitivity analysis are presented in Table S4 and Figure S1.

Table S1 MES-C parameters, values and units used in this study

Parameter	Description	Value	Units	Reference
V_{slope}	Regression coefficient (S4)	0.063	ln(mg C_s (mg MIC)$^{-1}$ h^{-1}) °C^{-1}	(Wieder et al., 2015)
V_{int}	Regression intercept (S4)	5.47	ln(mg C_s (mg MIC)$^{-1}$ h^{-1})	Same as above
a_v	Tuning coefficient (S4)	8×10^{-6}		Same as above
$V_{mod\text{-}r}$	Modifies V_{max} for fluxes into MIC_r (S4)	10,2,10[a]		Same as above
$V_{mod\text{-}K}$	Modifies V_{max} for fluxes into MIC_K (S4)	3,3,2[b]		Same as above
K_{slope}	Regression coefficient (S5)	0.017, 0.027, 0.017 [a,b]		Same as above
K_{int}	Regression intercept (S5)	3.19		Same as above
a_k	Tuning coefficient (S6)	10		Same as above
$K_{mod\text{-}r}$	Modifies Km for fluxes into MIC_r (S5)	$0.125,0.5,0.25\times P_{scalar}$ [a]		Same as above
$K_{mod\text{-}K}$	Modifies Km for fluxes into MIC_K (S5)	$0.5,0.25,0.167\times P_{scalar}$ [b]		Same as above
P_{scalar}	Physical protection scalar used in K_{mod}	$(2.0 \times e^{-2\sqrt{f_{clay}}})^{-1}$		Same as above
τ_r	r-selected microbial biomass turnover rate (S9)	$5.2 \times 10^{-4} \times e^{0.3(f_{met})} \times \tau_{mod}$	h^{-1}	Same as above
τ_k	K-selected microbial biomass turnover rate (S13)	$2.4 \times 10^{-4} \times e^{0.1(f_{met})} \times \tau_{mod}$		Same as above
τ_{mod}	Modifies microbial turnover rate	$0.6 < \sqrt{NPP/100} < 1.3$		This study
f_{met}	Partitioning of litter inputs to LIT_m	0.85-0.013 (lignin/N)		(Wieder et al., 2015)
k_{adsorp}	Binding affinity of LMWC (S14)	$1/24.0 \times f_{pH}$	h^{-1}	(Abramoff et al., 2022)
f_{pH}	pH regulation on binding affinity	$e^{-0.186\times pH-0.216}$		Same as above
Q_{max}	Maximum sorption capacity (S14)	$0.6 \times \%claysilt$ [c]	mg C (g soil)$^{-1}$	(Georgiou et al., 2022)
$K(\theta,\phi)$ [d]	Moisture scalar (S14)	$(\theta/\phi)^{0.5}$		(Ghezzehei et al., 2019)
k_{desorp}	Desorption rate (S15)	1/6.0/24.0	h^{-1}	(Wang et al., 2013)
k_{ab}	Breakdown rate of soil aggregate carbon (S16)	1.14×10^{-5}	h^{-1}	Same as above
k_{ma}	Rate of aggregate formation from MAOC (S19)	4.57×10^{-6}	h^{-1}	Same as above
k_l	Leaching rate of LMWC (S20)	0.0015/24	h^{-1}	(Abramoff et al., 2022)
f_m	Fraction of metabolic litter inputs transferred to AGG_m (S21)	0.5		This study

f_s	Fraction of structural litter inputs transferred to AGG_s (S22)	0.5		This study
MGE	Microbial growth efficiency (S23)	0.55,0.25,0.75,0.35×T_{scalar} [e]	mg mg^{-1}	(Wieder et al., 2015)
T_{scalar}	Temperature regulation on MGE	$e^{(-0.015\times fT)}$		This study
f_a	Fraction of τ partitioned to LWMC (S25)	1-f_c		(Wieder et al., 2015)
f_c	Fraction of τ partitioned to MAOC (S26)	$0.1 \times e^{-3(fmet)}$,$0.3 \times e^{-3(fmet)}$ [f]		Same as above

[a] For LIT_m, LIT_s and LMWC fluxes entering MIC_r, respectively.

[b] For LIT_m, LIT_s and LMWC fluxes entering MIC_K, respectively.

[c] %claysilt represents soil clay and silt content.

[d] θ is soil moisture (mm^3 mm^{-3}) and ϕ is soil porosity (mm^3 mm^{-3}).

[e] The first two values correspond to C fluxes into MIC_r, the second two values correspond to C fluxes into MIC_K.

[f] For MIC_r and MIC_K, respectively.

Table S2 The optimised MIMICS parameters (dimensionless) and their value ranges

Parameter	Definition	Range
xav	A scaling factor for V_{max}	0.1-10.0
xak	A scaling factor for K_m	0.5-20.0
xdesorp	A scaling factor for SOC desorption rate	0.1-2.0
xbeta	An exponent of the biomass-dependent mortality rate of microbes	1.0-2.0
xdiff	A scaling factor for SOC diffusion coefficient in soils	0.1-10.0
xnpp	A scaling factor for litter input represented by NPP	0.5-2.0

Table S3 The optimised MES-C parameters (dimensionless) and their value ranges

Parameter	Definition	Range
xak	A scaling factor for K_m	0.1-10.0
xfm	A scaling factor for metabolic litter inputs entering AGG_m	0.1-5.0
xtvp	A scaling factor for aggregated C breakdown rate	0.1-10.0
xtvc	A scaling factor for aggregation formation rate from MAOC	0.1-10.0
xqmax	A scaling factor for maximum adsorption capacity	0.5-5.0
xnpp	A scaling factor for litter input represented by NPP	0.5-2.0

Table S4 Optimised parameter values for vertically resolved MIMICS

	xav	xak	xdesorp	xbeta	xdiff	xnpp
Cluster 1	2.887	19.911	1.927	1.379	6.226	1.209
Cluster 2	1.629	16.437	1.412	1.060	0.456	0.518
Cluster 3	1.116	18.079	1.992	1.498	0.111	0.516
Cluster 4	2.078	12.819	1.041	1.269	9.975	0.853
Cluster 5	3.635	15.716	1.080	1.454	4.986	1.427
Cluster 6	3.556	19.770	1.950	1.319	9.592	1.117
Cluster 7	1.740	9.820	0.885	1.512	5.629	1.835
Cluster 8	3.408	19.662	1.961	1.260	6.108	1.377
Cluster 9	1.014	18.997	1.958	1.485	0.928	0.508
Cluster 10	2.215	15.570	1.910	1.675	6.095	1.770
Cluster 11	2.754	17.742	1.714	1.677	5.381	1.570
Cluster 12	1.650	13.848	1.928	1.515	2.988	1.862

Table S5 Optimised parameter values for vertically resolved MES-C

	xak	xfm	xtvp	xtvc	xqmax	xnpp
Cluster 1	1.746	4.480	6.050	0.265	0.542	0.541
Cluster 2	4.345	2.556	4.051	0.315	0.503	1.396
Cluster 3	0.532	1.197	5.365	3.492	0.643	1.752
Cluster 4	4.052	0.388	6.075	1,151	0.505	0.525
Cluster 5	0.268	4.806	8.971	2.828	0.604	1.026
Cluster 6	0.584	1.218	5.233	1.448	0.624	1.586
Cluster 7	0.780	4.875	5.736	0.662	0.587	1.301
Cluster 8	0.856	2.636	7.446	0.257	0.503	1.773
Cluster 9	1.460	3.027	7.004	1.195	0.511	1.338
Cluster 10	0.550	0.158	0.826	0.243	0.528	1.682
Cluster 11	0.553	4.714	6.736	0.326	0.656	0.929
Cluster 12	0.595	3.978	6.488	0.326	0.534	1.860

Table S6 Parameters for sensitivity analysis in MES-C and their ranges and representation in MES-C model. All parameters are unitless scaling factor.

Parameters	Representation in MES-C	Range
xav	Multiplied by a_v in Table S1	0.1-10.0
xak	Multiplied by a_k in Table S1	0.1-10.0
xfm	Multiplied by f_m in Table S1	0.1-5.0
xfs	Multiplied by f_s in Table S1	0.1-5.0
xtvmic	Multiplied by μ_{mic} in (S9) and (S13)	0.1-10.0
xtvp	Multiplied by k_{ab} in Table S1	0.1-10.0
xtvc	Multiplied by k_{ma} in Table S1	0.1-10.0
xtvac	Multiplied by k_l in Table S1	0.1-10.0
xkba	Multiplied by k_{desorp} in Table S1	0.5-2.0
xqmax	Multiplied by Q_{max} in Table S1	0.5-5.0
xdiffsoc	Multiplied by vertical diffusion coefficient in soil (default value 1.1×10^{-4} cm^2 $hour^{-1}$)	0.1-10.0
xnpp	Multiplied by NPP	0.5-2.0
xvmaxbeta	Multiplied by z_τ in (S2)	0.5-5.0

Table S7 Minimum and maximum performance metrics of SOC predictions for test data. Values in brackets are the averages of 10 cross-validation simulations. Unit for depth is cm, and unit for RMSE is log(g C kg-1 soil).

Model	Depth	RMSE	R^2	LCCC	AIC
RF_{env}	0-30	0.59-0.61 (0.59)	0.59-0.62 (0.61)	0.70-0.72 (0.71)	17253-17649 (17437)
	30-60	0.84-0.89 (0.86)	0.42-0.45 (0.44)	0.54-0.57 (0.55)	14035-14961 (14448)
	60-90	0.88-0.90 (0.89)	0.43-0.45 (0.44)	0.54-0.56 (0.55)	10347-11571 (10864)
	90-120	0.88-0.90 (0.89)	0.43-0.45 (0.44)	0.54-0.56 (0.55)	10347-11571 (10864)
	120-150	0.88-0.91 (0.89)	0.40-0.43 (0.42)	0.53-0.56 (0.54)	5190-6349 (5781)
RF_{inp}	0-30	0.72-0.75 (0.73)	0.39-0.42 (0.40)	0.51-0.53 (0.53)	18416-18870 (18650)
	30-60	0.97-1.01 (0.98)	0.26-0.29 (0.28)	0.37-0.40 (0.38)	15081-16134 (15512)
	60-90	1.01-1.03 (1.02)	0.27-0.29 (0.28)	0.37-0.39 (0.38)	11416-12403 (11750)
	90-120	1.00-1.03 (1.01)	0.25-0.29 (0.27)	0.36-0.40 (0.38)	8575-9305 (8896)
	120-150	0.98-1.01 (1.00)	0.27-0.33 (0.29)	0.39-0.43 (0.41)	5742-6818 (6294)
MIMICS	0-30	0.85-0.89 (0.86)	0.13-0.16 (0.15)	0.29-0.33 (0.32)	19315-19618 (19431)
	30-60	1.09-1.14 (1.11)	0.05-0.07 (0.06)	0.12-0.15 (0.14)	16053-16588 (16287)
	60-90	1.20-1.25 (1.22)	0.03-0.05 (0.04)	0.08-0.10 (0.09)	11948-12930 (12365)
	90-120	1.28-1.33 (1.30)	0.03-0.04 (0.03)	0.06-0.08 (0.07)	8501-9871 (9143)
	120-150	1.37-1.41 (1.39)	0.01-0.03 (0.02)	0.04-0.06 (0.05)	6082-6769 (6479)
MES-C	0-30	0.84-0.88 (0.86)	0.14-0.16 (0.15)	0.33-0.36 (0.35)	18513-19833 (19507)
	30-60	1.03-1.12 (1.08)	0.06-0.07 (0.07)	0.19-0.23 (0.21)	15607-16969 (16546)
	60-90	1.13-1.22 (1.19)	0.05-0.06 (0.06)	0.16-0.19 (0.18)	11979-13250 (12591)
	90-120	1.14-1.24 (1.21)	0.06-0.08 (0.06)	0.17-0.19 (0.18)	8659-10045 (9259)
	120-150	1.07-1.11 (1.09)	0.07-0.08 (0.07)	0.22-0.24 (0.23)	5837-6755 (6349)

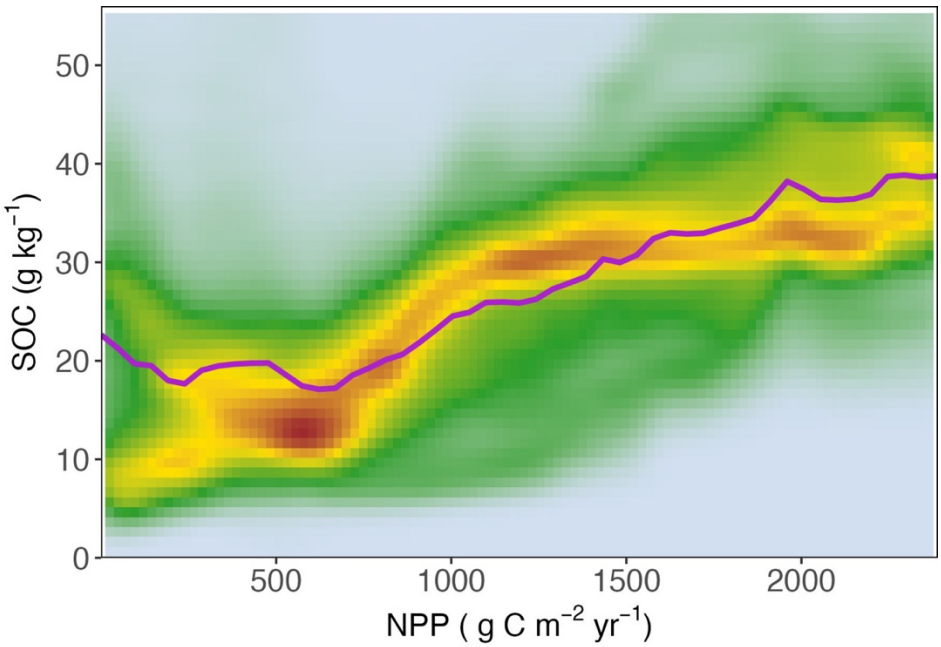

Figure S1. Density representation of ICE with respect to different predictors and observed SOC in natural ecosystems. The purple curves represent the mean partial dependence across all instances. Note that SOC greater than 56 g C kg^{-1} soil (around 1% of data) are not shown here for better visualisation of the dominant pattern.

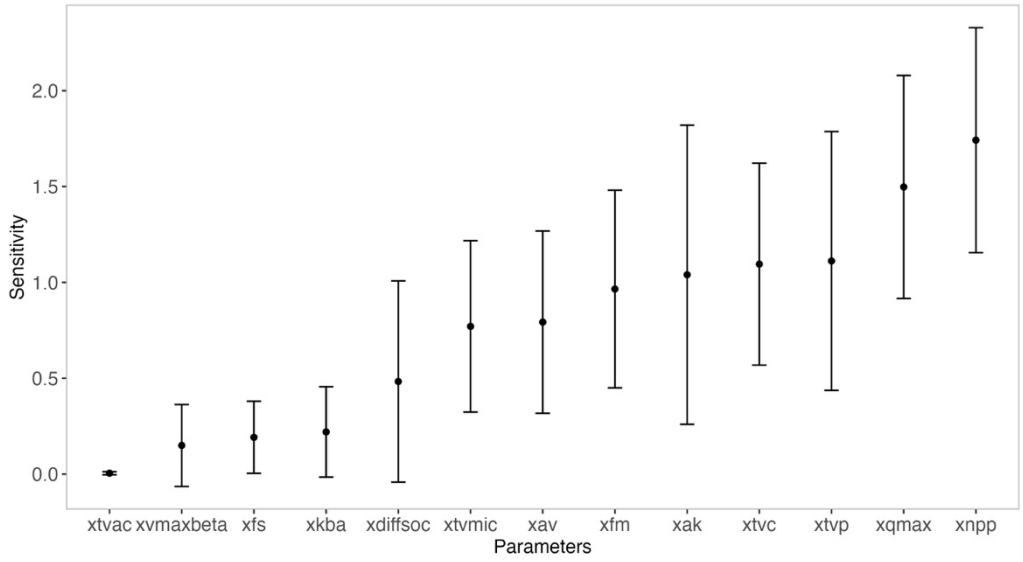

Figure S2. Mean parameter sensitivities (±1 SD) across 12 clusters in MES-C. Parameter definitions and ranges are provided in Table S6.

References

Abramoff, R. Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Rossel, R. A. V., Yuan, W. and Ciais, P. (2022). Improved global-scale predictions of soil carbon stocks with Millennial Version 2. *Soil Biology and Biochemistry* **164**: 108466.

Duan, Q., Gupta, V. K. and Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications* **76**: 501- 521.

Georgiou, K., Jackson, R. B., Vindušková, O., Abramoff, R. Z., Ahlström, A., Feng, W., Harden, J. W., Pellegrini, A. F., Polley, H. W. and Soong, J. L. (2022). Global stocks and capacity of mineral-associated soil organic carbon. *Nature Communications* **13**: 3797.

Ghezzehei, T. A., Sulman, B., Arnold, C. L., Bogie, N. A. and Berhe, A. A. (2019). On the role of soil water retention characteristic on aerobic microbial respiration. *Biogeosciences* **16**: 1187-1209.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E. and Schulze, E. D. (1996). A global analysis of root distribution for terrestrial biomes. *Oecologia* **108**: 389-411.

Koven, C. D., Riley, W. J., Subin, Z. M., Tang, J. Y., Torn, M. S., Collins, W. D., Bonan, G. B., Lawrence, D. M. and Swenson, S. C. (2013). The effect of vertically resolved soil biogeochemistry and alternate soil C and N models on C dynamics of CLM4. *Biogeosciences*, 10, 7109-7131, https://doi.org/10.5194/bg-10-7109-2013.

Lu, X., Wang, Y. P., Ziehn, T. and Dai, Y. (2013). An efficient method for global parameter sensitivity analysis and its application to the Australian community land surface model (CABLE). *Agriculture and forest meterorology* **182**: 292-303.

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

Wang, G., Post, W. M. and Mayes, M. A. (2013). Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. *Ecological Applications* **23**: 255-272.

Wang, Y. P., Zhang, H., Ciais, P., Goll, D., Huang, Y., Wood, J. D., Ollinger, S. V., Tang, X. and Prescher, A. K. (2021). Microbial activity and root carbon inputs are more important than soil carbon diffusion in simulating soil carbon profiles. *Journal of Geophysical Research: Biogeosciences* **126**, e2020JG006205.

Wieder, W., Grandy, A., Kallenbach, C., Taylor, P. and Bonan, G. (2015). Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development* **8**: 1789-1808.

Wieder, W. R., Sulman, B. N., Hartman, M. D., Koven, C. D. and Bradford, M. A. (2019). Arctic soil governs whether climate change drives global losses or gains in soil carbon. *Geophysical Research Letters* **46**: 14486-14495.