SUPPLEMENTARY MATERIALS FOR:

Unifying soil organic matter formation and persistence frameworks: the MEMS model

FULL MODEL DESCRIPTION OF MEMS V1.0

Mathematical representation of MEMS v1.0

Below are the differential equations for dynamics through time as calculated by MEMS v1.0. For simplicity, many of the individual fluxes are summarized by single names (e.g., $C1_{in}^{i}$ to represent total inputs to the C1 pool from litter material *i*, instead of including the separate calculation). Please refer to the equations provided in this Supplementary Materials. Parameter descriptions can be found in Table 2 of the main manuscript. Please note that the below list equations are fully representative of the carbon dynamics of MEMS v1.0 but are layer- and time-specific. However, for simplicity are presented in a generalized form.

$$\frac{dC1}{dt} = c1_{in}^{i} - (\boldsymbol{uk} * \boldsymbol{C1} * \boldsymbol{k_1}) \quad \textbf{(1)}$$

$$\frac{dC2}{dt} = c2_{in}^{i} - (\boldsymbol{uk} * \boldsymbol{C2} * \boldsymbol{k_2}) - (\boldsymbol{C2} * \boldsymbol{LIT_{frg}}) \quad \textbf{(2)}$$

$$\frac{dC3}{dt} = c3_{in}^{i} - (\boldsymbol{C3} * \boldsymbol{k_3}) - (\boldsymbol{C3} * \boldsymbol{LIT_{frg}}) \quad \textbf{(3)}$$

$$\frac{dC4}{dt} = c4_{ass}^{C1} + c4_{ass}^{C2} - (\boldsymbol{C4} * \boldsymbol{k_4}) \quad \textbf{(4)}$$

$$\frac{dC5}{dt} = c5_{gen}^{C4} + c5_{frg}^{C2} + c5_{frg}^{C3} - (\boldsymbol{C5} * \boldsymbol{k_5}) \quad \textbf{(5)}$$

$$\frac{dC6}{dt} = c6_{in}^{i} + c6_{in}^{C1} + c6_{in}^{C2} + c6_{in}^{C3} + c6_{in}^{C4} - c8_{in}^{C6} \quad \textbf{(6)}$$

$$\frac{dC7}{dt} = \boldsymbol{C1_{co2}} + \boldsymbol{C2_{co2}} + \boldsymbol{C3_{co2}} + \boldsymbol{C4_{co2}} + \boldsymbol{C5_{co2}} + \boldsymbol{C8_{co2}} + \boldsymbol{C9_{co2}} + \boldsymbol{C10_{co2}} \quad \textbf{(7)}$$

$$\frac{dC8}{dt} = c8_{in}^{C5} + c8_{in}^{C6} + c8_{in}^{C10} - \boldsymbol{sorption} - (\boldsymbol{C8} * \boldsymbol{DOC_{lch}}) - (\boldsymbol{C8} * \boldsymbol{k_8}) \quad \textbf{(8)}$$

$$\frac{dC9}{dt} = \boldsymbol{sorption} - (\boldsymbol{C9} * \boldsymbol{k_9}) \quad \textbf{(9)}$$

$$\frac{dC10}{dt} = c10_{frg}^{C2} + c10_{frg}^{C3} - (\boldsymbol{C10} * \boldsymbol{k_{10}}) \quad \textbf{(10)}$$

$$\frac{dC11}{dt} = (\boldsymbol{C8} * \boldsymbol{DOC_{lch}}) \quad \textbf{(11)}$$

Carbon inputs from external sources

In MEMS v1.0 the above- and below-ground plant residue inputs are combined and input to the system on a daily timestep. These total inputs are partitioned between C1, C2, C3 and C6 as a function of the external source (i) input properties (Eqs. 12-15): the cold water extractable fraction of the hot-water extractable litter input (f_{DOC}^{i}), the hot water extractable fraction of the litter input (f_{SOL}^{i}) and acid-insoluble fraction of the litter input (f_{LIG}^{i}).

$$ {}_{j}C1_{in}^{i} = ({}_{j}CT^{i} * f_{SOL}^{i}) - ({}_{j}CT^{i} * f_{SOL}^{i} * f_{DOC}^{i}) \quad (12) $$

$$ {}_{j}C2_{in}^{i} = {}_{j}CT^{i} - \left({}_{j}CT^{i} * (f_{SOL}^{i} + f_{LIG}^{i})\right) \quad (13) $$

$$ {}_{j}C3_{in}^{i} = ({}_{j}CT^{i} * f_{LIG}^{i}) \quad (14) $$

$$ {}_{j}C6_{in}^{i} = {}_{j}CT^{i} * f_{SOL}^{i} * f_{DOC}^{i} \quad (15) $$

Where ${}_{j}X_{in}^{i}$ is refers to the daily carbon input to pool X from external source i on day j, and ${}_{j}CT^{i}$ is the total daily carbon input from external source i on day j. For MEMS v1.0 the layer is fixed to the aboveground litter layer only, allowing for use of the same functions as those presenting in the LIDEL model (Campbell *et al*., 2016). However, future versions may incorporate the same structure for different points of entry for C inputs (e.g., root death and the rhizosphere).

Once allocated to their initial pools, the carbon is susceptible to assimilation in microbial biomass if it is water-soluble (C1) or acid-soluble (C2) but only co-metabolized if it is acid-insoluble (C3). The contents of these pools represent compounds of increasing chemical complexity (e.g., C1, mostly soluble carbohydrates, phenols and amino acids; C2, mostly cellulose, xylans and other hemicelluloses; C3, mostly lignin aboveground and suberin/cutin belowground) and are associated with decreasing microbial use efficiency.

Microbial assimilation from litter pools

Many of the biogeochemical processes represented by MEMS are assumed to be microbially mediated, and therefore are associated with C-mineralization and the resulting carbon dioxide (CO_2) emissions from microbial respiration. The primary carbon losses result from the metabolic processes of bacteria and fungi within the soil and are aligned with the mathematical representations as described by Campbell *et al*. (2016) and, in part, summarise the findings of Sinsabaugh *et al*. (2013), Moorhead *et al*. (2013) and Soong *et al*. (2015). In addition, carbon assimilation by microbial biomass (C4) in the litter layer results from the balance between anabolic and catabolic processes and thus, as biomass is formed, dissolved organic matter (DOM) and CO_2 are also produced . Microbial assimilation is a function of nitrogen content and lignocellulosic index (Eq. 16) of the structural litter pools (C2 and C3; organic matter > 2 mm) and controlled by maximum decomposition rates for C1 (k_1) and C2 (k_2) that assume first-order decay.

$$ {}_jLCI_{lit} = \frac{{}_jC3}{\left({}_jC2 + {}_jC3\right)} \quad (16) $$

$$ {}_jC4_{ass}^{C1} = uB * B_1 * (1 - la_4) * uk * k_1 * {}_jC1 \quad (17) $$

$$ {}_jC4_{ass}^{C2} = uB * B_2 * (1 - la_1) * uk * k_2 * {}_jC2 \quad (18) $$

Where ${}_jC4_{ass}^{C1}$ and ${}_jC4_{ass}^{C2}$ refer to the fraction of the given litter pool (i.e., C1 or C2) that is microbially assimilated to pool C4 on day *j* from pool C1 or C2, respectively. Note that these functions are make microbial assimilation explicit in this aboveground litter layer. In the soil itself, microbial assimilation of organic matter is still occurring but assumed to be implicit and incorporated in the carbon mineralization rates for each of the soil pools (e.g., C5, C8, C9 and C10). In future versions of the model, the same general structure can apply, with an explicit microbial component at the different points of entry (i.e., rhizospheric inputs vs aboveground

litter) but parameter values may differ between layers, when more are added. Detail about the concepts behind this approach can be found in Sokol *et al*., 2018.

More information of the parameters uB, uk, B_x, la_x and k_x can be found in Campbell *et al*. (2016) and in the equations below, but briefly:

- ${}_ju B$ and ${}_juk$ are rate modifiers to represent the litter chemistry controls (LCI and available nitrogen) on microbial use efficiency, on day *j*.

$$ {}_j\boldsymbol{uB} = \boldsymbol{min}\left(\left(\frac{1}{1+e^{-N_{max}(\,N_{lit}-N_{mid})}}\right),\left(1-e^{-0.7\left(\left|\,{}_jLCI_{lit}-0.7\right|*10\right)}\right)\right) \quad \textbf{(19)} $$

$$ {}_j\boldsymbol{uk} = \boldsymbol{min}\left(\left(\frac{1}{1+e^{-N_{max}(\,N_{lit}-N_{mid})}}\right),\left(e^{-3*\,{}_jLCI_{lit}}\right)\right) \quad \textbf{(20)} $$

Where N_{max} and N_{mid} are maximum and mid points of litter nitrogen content having an impact on microbial use efficiencies, using a logistic curve (see Figure S7). N_{lit} and ${}_jLCI_{lit}$ are the input material nitrogen content and LCI being simulated on day *j*.

IMPORTANT NOTE – In MEMS v1.0 there is no nitrogen cycling and therefore the N_{lit} value is not dynamic, as it likely should be. Consequently, MEMS v1.0 uses the nitrogen content of the input material, and therefore N_{lit} is a constant through time and across layers. This constant nitrogen value is consistent with the approach used by the LIDEL model (Campbell *et al*., 2016) however it is expected that a dynamic nitrogen (i.e. be ${}_jN_{lit}$ – as equivalent to ${}_jLCI_{lit}$) content would more likely reflect real-world conditions, especially in extended periods without litter input.

- B_1 and B_2 are maximum growth efficiencies associated with the water-soluble and acid-soluble litter pools (C1 and C2), respectively (See Table 2 in the main manuscript).

- la_1 and la_4 are estimates of carbon in DOM generation from leaching the decayed litter pools on day *j*.

$$ {}_{j}\boldsymbol{la_1} = \boldsymbol{min}\left(\left(\boldsymbol{E_{Hmax}} - \frac{(\boldsymbol{E_{Hmax}} - \boldsymbol{E_{Hmin}})}{\boldsymbol{LCI_{max}}} * {}_{j}\boldsymbol{LCI_{lit}}\right), \left(\boldsymbol{E_{Hmax}} - \frac{(\boldsymbol{E_{Hmax}} - \boldsymbol{E_{Hmin}})}{\boldsymbol{N_{max}}} * \boldsymbol{N_{lit}}\right)\right) \quad \textbf{(21)} $$

$$ {}_{j}\boldsymbol{la_4} = \boldsymbol{min}\left(\left(\boldsymbol{E_{Smax}} - \frac{(\boldsymbol{E_{Smax}} - \boldsymbol{E_{Smin}})}{\boldsymbol{LCI_{max}}} * {}_{j}\boldsymbol{LCI_{lit}}\right), \left(\boldsymbol{E_{Smax}} - \frac{(\boldsymbol{E_{Smax}} - \boldsymbol{E_{Smin}})}{\boldsymbol{N_{max}}} * \boldsymbol{N_{lit}}\right)\right) \quad \textbf{(22)} $$

Where E_{Hmax} and E_{Hmin} are the maximum and minimum amount of DOM leached from decay of acid-soluble litter (C2), and E_{Smax} and E_{Smin} are the maximum and minimum amount of DOM leached from decay of water-soluble litter (C1). LCI_{max} refers to the maximum lignocellulosic index that can have an impact on these rates. As noted above, N_{lit} and ${}_{j}LCI_{lit}$ are the nitrogen content of input material and LCI being simulated on day *j*.

- k_1 and k_2 are the maximum decay rates of water-soluble (C1) and acid-soluble (C2) litter pools, respectively (See Table 2 in the main manuscript).

Microbial mortality and necromass production

After carbon is metabolized by microbes and incorporated in pool C4, the death and products of microbial activity result in the compounds that form the coarse, heavy particulate SOM (C5) that is often found coating sand particles in the > 53 µm soil fraction (Ludwig *et al*., 2015). In the aboveground litter layer simulated by MEMS v1.0, this process of microbial biomass decay results in loss to DOC (C6) and CO_2 (C7), in addition to the C5 pool belowground.

$$ {}_{j}\boldsymbol{C5_{gen}^{C4}} = \boldsymbol{B_3} * (\boldsymbol{1} - \boldsymbol{la_2}) * \boldsymbol{k_4} * {}_{j}\boldsymbol{C4} \quad \textbf{(23)} $$

Where ${}_{j}C5_{gen}^{C4}$ refers to the fraction of carbon that is transferred from C4 to C5 (i.e., microbial products transported belowground when physical and hydrological processes mix between the

input layer [aboveground litter only in MEMS v1.0] and soil layer) on day j. The flux from the aboveground microbial biomass pool (C4) is assumed to move belowground, to the first soil layer (see Figure 1 in the main manuscript). More information of the parameters B_3, la_2 and k_4 can be found in Table 2 in the main manuscript, but briefly, B_3 refers to a maximum rate of microbial product (C5) generation per unit of microbial biomass (C4) decayed, la_2 refers to the maximum amount of DOM produced per unit of microbial biomass (C4) decayed and k_4 refers to the maximum rate of microbial biomass (C4) decay.

Fragmentation and perturbation

To quantify the transfer of carbon from large (> 2 mm) particulates to small particulates belowground, simple parameter values have been allocated to represent first-order rates of transfer from both structural litter pools (C2 and C3). As model development continues, these rates will be improved to provide more mechanistic relationships with site conditions (see Braakehekke *et al*., 2011). See Table 2 for information about the parameter used in MEMS v1.0 (LIT_{frg}). The amount of litter C fragmented and transferred vertically from structural litter pools to the belowground POM pools (C5 and C10) is also governed by the POM_{split} parameter that defines how much of the total is allocated to C5.

$$ {}_jC5_{frg}^{C2} = POM_{split} * LIT_{frg} * {}_jC2 \quad (24) $$

$$ {}_jC5_{frg}^{C3} = POM_{split} * LIT_{frg} * {}_jC3 \quad (25) $$

$$ {}_jC10_{frg}^{C2} = \left(1 - POM_{split}\right) * LIT_{frg} * {}_jC2 \quad (26) $$

$$ {}_jC10_{frg}^{C3} = \left(1 - POM_{split}\right) * LIT_{frg} * {}_jC3 \quad (27) $$

Where ${}_jCX_{frg}^{CY}$ refers to the amount of carbon that is transferred from pool *CY* to pool *CX* on day *j*.

Dissolved organic matter production

Dissolved organic matter plays a major role in the MEMS model as it is the only way in which carbon can sorb to mineral surfaces in the soil, meaning that if there is limited DOM there will also be limited stabilization in MAOM (C9). Consequently, DOM production from all model pools is simulated explicitly according to the formulae provided by the LIDEL model (Campbell *et al*., 2016) and based on empirical data in Soong *et al*. (2015). Each timestep, the aboveground litter layer DOM (C6) receives a fraction of inputs from external sources directly (Eq. 15; ${}_jC6_{in}^{i}$), from all litter layer pools (${}_jC6_{in}^{C1}$, ${}_jC6_{in}^{C2}$, ${}_jC6_{in}^{C3}$) and from microbial biomass (${}_jC6_{in}^{C4}$).

$$ {}_j\boldsymbol{C6_{in}^{C1}} = \boldsymbol{la_4 * uk * k_1 * \ {}_jC1} \quad \textbf{(28)} $$

$$ {}_j\boldsymbol{C6_{in}^{C2}} = \boldsymbol{la_1 * uk * k_2 * \ {}_jC2} \quad \textbf{(29)} $$

$$ {}_j\boldsymbol{C6_{in}^{C3}} = \boldsymbol{la_3 * k_3 * \ {}_jC3} \quad \textbf{(30)} $$

$$ {}_j\boldsymbol{C6_{in}^{C4}} = \boldsymbol{la_2 * k_4 * \ {}_jC4} \quad \textbf{(31)} $$

Where ${}_jCx_{in}^{Cy}$ refers to DOM leaching from pool *y* to pool *x* on day *j*. The parameters used are detailed in Table 2 in the main manuscript, and/or defined in previous equation in this section. Note that pool C6 is not the DOM consumed by microbial biomass but rather the amount leftover after microbial activity. In this initial model version, the litter layer only refers to the aboveground component, but the same structure can equally apply to belowground C inputs such as root death. However, measurably, the DOM in the C6 pool is directly equivalent to the belowground soil DOM (C8). In MEMS v1.0, DOM enters the soil through the C6 pool only. When explicit inputs from belowground litter (e.g., roots) are simulated in future versions Eqs. 28-31 can apply for each

soil layer adding the DOM that is in excess of microbial activity directly to pool C8 instead of the 'C6' shown in the equations above. Similarly, root exudates can be simulated as direct addition to the C8 pool of any specific soil layer. Hence, just as the litter layer DOM (C6) receives inputs from the aboveground litter layer pools, the soil DOM (C8) would receive inputs from the belowground pools (e.g., decomposing root matter and root exudation). In addition, the soil DOM pool receives inputs from the POM and MAOM pools (${}_jC8_{in}^{C5}$, ${}_jsorption$, ${}_jC8_{in}^{C10}$) as well as from leached litter DOM (C6). Here, the $sorption$ flux represents the net carbon exchange between soil DOM (C8) and MAOM (C9).

$$ {}_jC8_{in}^{C5} = la_3 * k_5 * {}_jC5 \quad (32) $$

$$ {}_jC8_{in}^{C6} = DOC_{frg} * {}_jC6 \quad (33) $$

$$ {}_jC8_{in}^{C10} = la_3 * k_{10} * {}_jC10 \quad (34) $$

The parameter values are defined in Table 2 in the main manuscript. As with the LIT_{frg} parameter, the DOC_{frg} value in MEMS v1.0 is set as a tuning parameter and simply assumes first-order rates to allocate a given proportion of the carbon in litter layer DOM pool (C6) to the soil DOM pool (C8) each timestep. As noted earlier, these functions are layer-specific and therefore in a multi-layer version of MEMS, there would be vertical leaching of DOM between C8 pool of different layers, instead of from the aboveground C6 pool alone (i.e., to replace Eq. 33).

Sorption and desorption

The formation of organo-mineral complexes in MEMS v1.0 is represented by a net sorption-desorption process that uses the amount of soil DOM (C8) to estimate adsorption rates based on a Langmuir isotherm (Kothawala *et al*., 2008). The key elements of this isotherm are the 'binding affinity' (K_{lm}) – see Eq. 35 – and maximum sorption capacity (Q_{max}) – see Eq. 36 – which are

controlled by site-specific conditions (soil pH and soil texture, respectively). It is worth noting that each of these site-specific conditions are provided as driving variables to the model, and are constants that represent the site at time-zero (i.e., soil pH is not simulated to change through time). The net sorption rate ($sorption$) aims to account for several different sorption mechanisms (e.g., cation bridging, surface complexation, etc.) to retain parsimony. A more accurate net flux may simulate the different mechanisms individually to allow for more detailed representation of different mineralogies as per Six *et al*. (2002) (e.g., dominated by 2:1 clays *vs* 1:1 clays). Future development of MEMS may adopt these changes.

$$\boldsymbol{K_{lm} = 10^{(-0.186 \quad soilpH - 0.216)}} \tag{35}$$

Where $soilpH$ refers to the 'native' soil pH of the simulated soil. The soil pH, as used in Eq 35, acts as a proxy for mineralogical differences between soils, with higher native soil pH being equated with weaker chemical bonding. This tenet is adopted from the regression provided in Mayes *et al*. (2012) and results in K_{lm} being estimated as in the MILLENNIAL model (Abramoff *et al*., 2017). However, the MEMS v1.0 estimate of Q_{max} does not follow the MILLENNIAL model and instead calculates a general relationship between maximum soil carbon capacity and soil texture using the entire dataset of Six *et al*. (2002). This takes a simple linear regression approach using the soil layer's percent silt and clay content (i.e., $100 - sand$)

$$\boldsymbol{Q_{max} = \quad \rho * \left(0.26126 * \left(100 - \quad sand\right) + 11.07820\right) * \left(1 - \quad rock\right)} \tag{36}$$

Where ρ refers to the bulk density of the soil at the site being simulated. Note that the bulk density is a conversion specific to the depth of the soil layer that converts a concentration from the regression of Six *et al*. (2002) to carbon density (e.g., gC m^{-2} layer $depth^{-1}$) and therefore the equations shown here assume a 1 meter deep layer for simplification. Both the sand content ($sand$) and rock fraction ($rock$) are expressed in percent (i.e., 0-100). The resulting equation

to represent net sorption is controlled by a Langmuir saturation function, using the amount of soil DOC (C8) available for sorption as well as the saturation deficit of MAOM (C9). Note, all coefficients in the equation below are layer- and timestep-specific.

$$ {}_{j}\boldsymbol{sorption} = {}_{j}\boldsymbol{C8} * \frac{\left(\left(\frac{\left(K_{lm} * Qmax * {}_{j}C8\right)}{1+\left({}_{j}K_{lm} * {}_{j}C8\right)}\right) - {}_{j}C9\right)}{Q_{max}} \quad \textbf{(37)} $$

Where ${}_{j}sorption$ is a net exchange of carbon between the soil DOM (C8) and MAOM (C9) pools given their size on day j. Since K_{lm} and Q_{max} are site-specific parameters, and the pool sizes (C8 and C9) are dynamic through time, there are interactions between these factors which mean sorption rates are not necessarily comparable between sites. This sorption process is assumed to be abiotic in that it results in no CO_2 emitted. As a net rate, sorption and desorption are not simulated individually which may make it difficult to represent potential priming effects on organo-mineral associations (e.g., Keiluweit *et al.*, 2015). Future MEMS model version will explore these feedbacks further.

Decomposition and pool decay rates

Apart from the litter layer DOM (C6), each of the state variables in MEMS v1.0 decay directly with unique decay rates informed by literature values (see Table 2). This decay results in CO_2 emissions which continually accumulate in the sink C7. The amount of CO_2 associated with each microbial process is equivalent to the amount of carbon leftover after losses to DOM are calculated so the decay rate constants for pool x (k_x) also embody explicit DOM generation and not just CO_2 emissions, as is more common in traditional SOM models (e.g., CENTURY or RothC). As with earlier equations, these below can be layer- and time-specific but for simplicity are presented in a generalized form.

$$ \boldsymbol{C1_{co2}} = \left(\left(\mathbf{1} - (\boldsymbol{uB} * \boldsymbol{B_1})\right) * (\mathbf{1} - \boldsymbol{la_4})\right) * \boldsymbol{uk} * \boldsymbol{k_1} * \boldsymbol{C1} \quad \textbf{(38)} $$

$$\boldsymbol{C2_{co2}} = \left(\left(\mathbf{1} - (\boldsymbol{uB} * \boldsymbol{B_2})\right) * (\mathbf{1} - \boldsymbol{la_1})\right) * \boldsymbol{uk} * \boldsymbol{k_2} * \boldsymbol{C2} \quad \textbf{(39)}$$

$$\boldsymbol{C3_{co2}} = (\mathbf{1} - \boldsymbol{la_3}) * \boldsymbol{k_3} * \boldsymbol{C3} \quad \textbf{(40)}$$

$$\boldsymbol{C4_{co2}} = \left((\mathbf{1} - \boldsymbol{B_3}) * (\mathbf{1} - \boldsymbol{la_2})\right) * \boldsymbol{k_4} * \boldsymbol{C4} \quad \textbf{(41)}$$

$$\boldsymbol{C5_{co2}} = (\mathbf{1} - \boldsymbol{la_3}) * \boldsymbol{k_5} * \boldsymbol{C5} \quad \textbf{(42)}$$

$$\boldsymbol{C8_{co2}} = \boldsymbol{k_8} * \boldsymbol{C8} \quad \textbf{(43)}$$

$$\boldsymbol{C9_{co2}} = \boldsymbol{k_9} * \boldsymbol{C9} \quad \textbf{(44)}$$

$$\boldsymbol{C10_{co2}} = (\mathbf{1} - \boldsymbol{la_3}) * \boldsymbol{k_3} * \boldsymbol{C10} \quad \textbf{(45)}$$

Where all parameters are defined in Table 2 in the main manuscript and earlier in this section. While the maximum decay rates (k_x) for most pools are fixed constants, Campbell *et al*. (2016) suggested that k_3 is best estimated in relation to the maximum decay rate of the microbially-accessible litter (C2) pool (k_2).

$$_j\boldsymbol{k_3} = \boldsymbol{k_2} * \left(\frac{\mathbf{0.2}}{\mathbf{1}+\frac{\mathbf{200}}{e^{\mathbf{8.15}*\ _j\boldsymbol{LCI_{lit}}}}}\right) \quad \textbf{(46)}$$

Note that when k_2 is a fixed value, k_3 only fluctuates with changes in the LCI of the litter layer. Also note that because the maximum decay rate of acid-insoluble litter (k_3) is determined relative to the LCI of all litter pools on a given day (j) the parameter itself can also be layer- and time-specific. At present, CO_2 emitted from soil DOM (determined by the maximum decay rate, k_8) is associated with the values presented in Kalbitz *et al*. (2005).

$$\boldsymbol{k_8} = \frac{\left(\left((\mathbf{0.000099})*\left(\frac{\mathbf{1}}{\mathbf{100}}\right)\right)+\left((\mathbf{0.000855})*\left(\frac{\mathbf{1}}{\mathbf{42}}\right)\right)+\left((\mathbf{0.001796})*\left(\frac{\mathbf{1}}{\mathbf{13}}\right)\right)\right)}{\boldsymbol{sum}\left(\left(\frac{\mathbf{1}}{\mathbf{100}}\right),\left(\frac{\mathbf{1}}{\mathbf{42}}\right),\left(\frac{\mathbf{1}}{\mathbf{13}}\right)\right)} \quad \textbf{(47)}$$

Decay rate modifiers

Soil temperature is simulated to have a polynomial relationship with decomposition, modifying each pool's decay rate according to the mean soil temperature of that layer on that day. The rationale behind this is to attempt to capture microbial processes and equate with realistic changes in enzymatic activity to be consistent with Michaelis-Menten kinetics. This follows the same function that is used by the STANDCARB 2.0 model (Harmon and Domingo, 2001) and produces a multiplier based on provided coefficients of optimum decomposition temperature (T_{opt}), the rate at which the decomposition rate increases with a 10 °C increase (T_{Q10}), the reference temperature at which that Q_{10} value was derived (T_{ref}), the shape of the excessive temperature limitation (T_{shp}) and the difference between optimum temperature and the decline above that threshold (T_{lag}).

$$ {}_{j}\boldsymbol{T_{mod}} = \boldsymbol{e}^{\left(-\left(\frac{{}_{j}soilT}{T_{opt}+T_{lag}}\right)\right)^{T_{shp}}} * \boldsymbol{T_{Q10}}^{\frac{{}_{j}soilT-T_{ref}}{T_{ref}}} \quad \textbf{(48)} $$

Where ${}_{j}T_{mod}$ is the temperature multiplier applied to decomposition of pools on day *j*, given the soil temperature on that day (${}_{j}soilT$). An initial MEMS v1.0 evaluation (prior to use with the LUCAS sites reported in the main manuscript), indicated the model consistently overestimated decomposition due to the temperature modifier effect. Consequently, the coefficients reported in Harmon and Domingo (2001) were revised down from those reported in Table 2 of the main manuscript (T_{opt} reduced to 35 °C, T_{shp} reduced to 3, T_{lag} increased to 7 °C and T_{Q10} increased to 3). In MEMS v1.0 this single function is used for all pools and over the single soil layer, however, it is also sufficiently generalizable to represent varying temperature sensitivities of the different pools (i.e., through the T_{Q10} coefficient) and of different layers. In which case, the temperature modifier would be specific to pool *x* on day *j* – e.g. ${}_{j}T^{x}_{mod}$. Furthermore, in future

versions of the MEMS model, we expect more explicit and complex relationships to temperature and moisture.

DOM transfer through soil layers

MEMS v1.0 does not have an explicit hydrological model, however this is likely needed for MEMS outputs to be reliably compared with empirical data at most sites (soil moisture often has a considerable influence on SOM formation and decomposition rates). Consequently, this is one of the first developments intended for MEMS. As a placeholder, leaching is assumed to be a unidirectional process with DOM lost to deeper soil layers (in the single-layer version) at a given maximum rate. This follows a first order rate of loss and simply assumes half the highest literature value found when performing a search of relevant studies.

DRIVING VARIABLES AND INITIALIZING MEMS V1.0

Site inputs and interpolating daily values from annual measurements

Driving variables of MEMS v1.0 can be either provided manually if they are known, or interpolated/estimated using basic site information. The format of this input information is typically in comma separated values (CSV) or any other ASCII text format and in R (R Core Team, 2018) is stored as a dataframe. As a single-layer, carbon model that only simulates litter and soil components of a site, MEMS v1.0 includes only a few *essential* driving variables. These fall into three major categories (climatic, edaphic and land use). For convenience, a summary of these essential inputs is provided in Table 3 of the main manuscript. The model operates on the assumption that a user must have measurements of soil pH, soil bulk density, annual NPP, sand content and rock fraction in order to simulate the site. Additionally, if daily temperature data are not known, the maximum, minimum and mean annual temperature can be used to interpolate daily values.

At the time of writing, daily soil temperature is the only climatic variable simulated in MEMS v1.0. The model can either be initialized using real, site-specific temperature data (if available), or daily values can be roughly estimated using a simple sine function related to the mean annual temperature (MAT) of the site (Eq. 49). This sine function provides 365 days of temperature values that are normally distributed around the MAT (therefore ensuring that the average from these daily values will also equal the MAT provided), with the peak of this sine on Julian day 182 (July 1st). This assumes the site is in the northern hemisphere but simulating a site in the southern hemisphere simply requires changing the sign of the 1.5 coefficient in Equation 49 below.

$$_{j}\boldsymbol{soilT} = \frac{\boldsymbol{T_{range}}}{\mathbf{2}} * \boldsymbol{sin}\big((\mathbf{2} * \boldsymbol{PIseq}) - \mathbf{1.5}\big) + \boldsymbol{MAT} \quad \textbf{(49)}$$

Where $_{j}soilT$ is the soil temperature in degrees Celsius on day j, T_{range} is the difference between the maximum daily soil temperature and minimum daily soil temperature measured over a year in degrees Celsius, PIseq is a sequence of 365 values evenly distributed from 0 to pi (≈ 3.14159), and MAT is the mean annual temperature in degrees Celsius of the site in question. While this approximation provides more realistic inputs than a constant temperature for each day, where possible, real, measured values should be imported separately as a list of average daily soil temperature values.

It should be noted that this sine function (with an intra-annual variation of T_{range} degrees Celsius) may not work well for sites near the equator where reduced seasonal dynamics mean that a smoothed sine curve does not represent reality. The T_{range} coefficient in Equation 49 is ideally calculated from estimates/measurements of a site's maximum and minimum soil temperatures of an average year, included alongside the MAT as inputs. However, these are optional and instead, a constant T_{range} value (i.e., the same range at all sites simulated) can be set as a global parameter

as shown in Table 2 in the main manuscript. This should be chosen carefully by the model user to best represent their site(s). It should also be noted that when simulating deeper soil layers they are also less likely to see large fluctuations in soil temperature and this should be considered when the user initializes multi-layer versions of the MEMS model.

Land use and management conditions

As with the sine function estimate soil temperature, the daily carbon inputs ($_jCT^i$) can also be estimated crudely according to a simplistic relationship with annual net primary productivity (NPP) – Equation 50).

$$_jCT^i = dnorm(seqDAY, peakDAY, sdNPP) * annNPP \quad (50)$$

Where $_jCT^i$ are the daily total carbon inputs from material i on day j, $seqDAY$ is a list of 365 integers that represent each day of the year, $peakDAY$ is a parameter value to specify the julian day of year when inputs peak (around which a normal distribution is generated) and $sdNPP$ is the 'width' of the distribution around the peak value. The $annNPP$ value is the site-specific annual NPP value in gC m^{-2} yr^{-1}. The $sdNPP$ parameter (specified as a global parameter) can be modified to represent different intra-annual distributions of the total carbon inputs. Specifically, this can change how 'quickly' the inputs are added to the soil (is the whole carbon input added within a few days or is it spread out over months?). For different land uses, $sdNPP$ may change according to the trends in plant growth at a given site. However, when simulating an equilibrium scenario where steady-state inputs are assumed, this has little or no effect over long simulations (i.e., 500+ years).

In most systems the total annual NPP is not directly equivalent to the total carbon inputs to the topsoil layer. Consequently, MEMS v1.0 reduces the annual amount based on how much of the

total can be realistically expected to be input to the specific layer given that site's land use. For example, Bolinder *et al*. (2007) suggest that, in arable sites where all residues are returned to soil, the proportion of annual NPP that is input to all soil varies between 55% and 78%. Whereas when all residues are removed, the proportion input can be as little as 21%. Furthermore, not all of this will be input to the topsoil layer simulated by MEMS v1.0. Consequently, before the daily inputs are interpolated from an annual value using Equation 50, the total is reduced based on best estimates for the land use and management routines of the site simulated.

$$ {}_{j}\boldsymbol{aCT^{i}} = {}_{j}\boldsymbol{CT^{i}} * \left(\frac{\mathbf{1}}{\boldsymbol{RtoS^{i}}+\mathbf{1}}\right) * (\mathbf{1} - {}_{j}\boldsymbol{aHARV^{i}}) \quad (51) $$

$$ {}_{j}\boldsymbol{bCT^{i}} = {}_{j}\boldsymbol{CT^{i}} * \left(\frac{\boldsymbol{RtoS^{i}}}{\boldsymbol{RtoS^{i}}+\mathbf{1}}\right) * (\mathbf{1} - {}_{j}^{L}\boldsymbol{bHARV^{i}}) \quad (52) $$

Where ${}_{j}aCT^{i}$ and ${}_{j}bCT^{i}$ are the aboveground and belowground carbon inputs of material *i* on day *j*. The aboveground and belowground split is achieved by use of a land-use specific root to shoot ratio of material *i* ($RtoS^{i}$) which are then reduced by fixed fractions (i.e., 0-1) to represent any losses through harvesting. Another parameter to describe natural losses due to weather (e.g., high winds) is also possible and resides as a placeholder in the general crop parameters file of MEMS v1.0. After the realistic aboveground fraction of NPP is derived, it can then replace the ${}_{j}CT^{i}$ term in Equation 50 and be used to interpolate daily inputs. However, the belowground fractions of NPP also includes inputs that are likely allocated to deeper soil layers than the topsoil simulated by MEMS v1.0. Consequently, the ${}_{j}bCT^{i}$ as calculated in Equation 52 is reduced by use of a Michaelis-Menten style function (see Kätterer *et al*., 2011) to proportion roots to the simulated soil layer.

$$ {}_{j}\boldsymbol{bCT^{i}} = {}_{j}\boldsymbol{bCT^{i}} * \left(\frac{\boldsymbol{depth}*(\boldsymbol{Rdep_{50}}+\boldsymbol{Rdep_{max}})}{\boldsymbol{Rdep_{max}}*(\boldsymbol{Rdep_{50}}+\ \boldsymbol{depth})}\right) \quad (53) $$

Where ${}_jbCT^i$ is the belowground carbon input of material i on day j, $depth$ is the depth of the soil in centimetres, $Rdep_{50}$ is the soil depth from the surface at which 50 % of the root biomass is proportioned in centimeters, and $Rdep_{max}$ is the maximum rooting depth in centimeters. These last two parameters are site specific but can be generalized according to different land-uses, reducing the number of inputs required by the model user. For information regarding these generalized parameters, see Canadell *et al*. (1996) and Jackson *et al*. (1996). For an example implementation of Equation 53 for the purpose of simulating SOM dynamics, see Poeplau (2016).

As with the interpolation of daily soil temperature from MAT, estimating daily values of carbon input are less precise than using real measured data. When possible, empirical data should be preferred and can be input along with daily climate data.

SUPPLEMENTARY FIGURES

(see attached files for high-resolution versions)

Figure S1 – Site information of all 8192 forest and grassland sites of the LUCAS dataset (Toth *et al*., 2013) used for validation of the MEMS v1.0 soil organic matter model. Different shapes represent different land use classes and all are overlaid over each other (grass = circles, n = 3487; broadleaved forests = triangle, n = 1590; mixed forest = crosses, n = 1402; coniferous forest = squares, n = 1713).

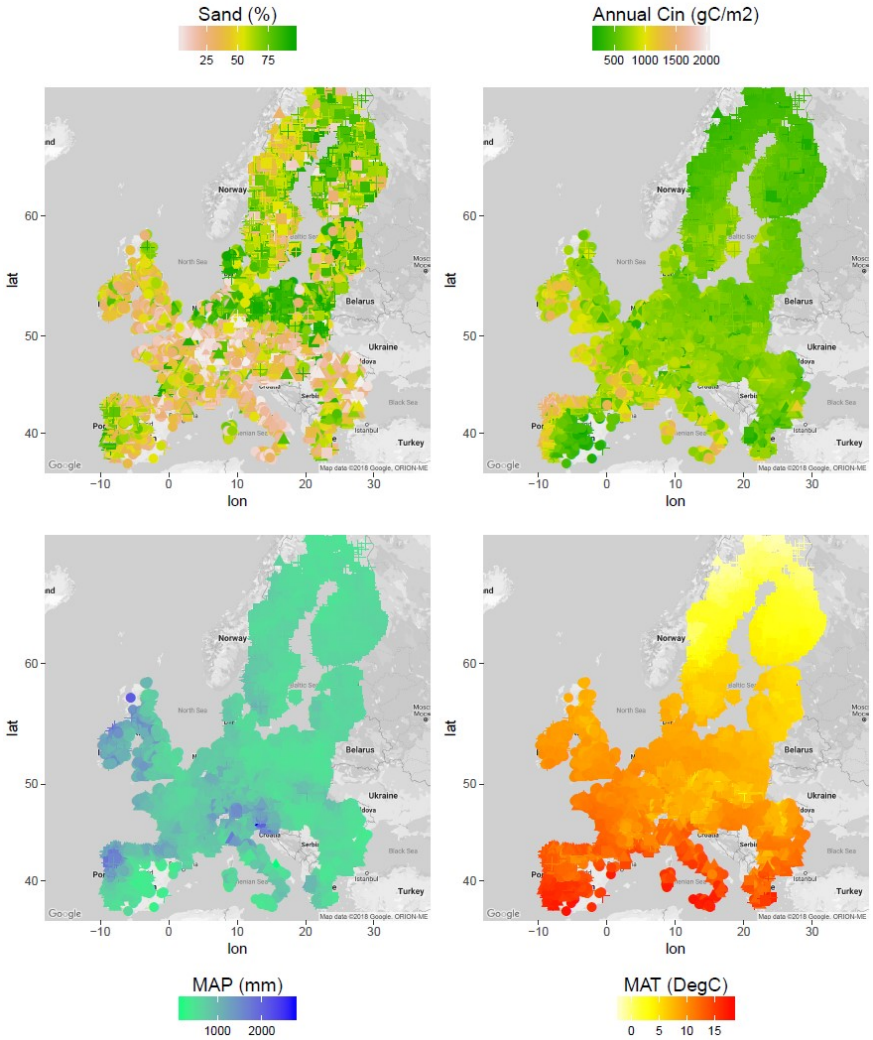


Figure S2 - Geographical distribution of 154 grassland and forest sites chosen for fractionation (a representative subsample of the total LUCAS database, see Toth *et al*., 2013). Reported mean annual temperature, mean annual precipitation and sand content are indicated for each site along with Net Primary Productivity (NPP) in 2009 derived from MODIS. Symbols indicate the land use division within grassland and forest. Cin is the C input, MAP is the mean annual precipitation and MAT is the mean annual temperature.

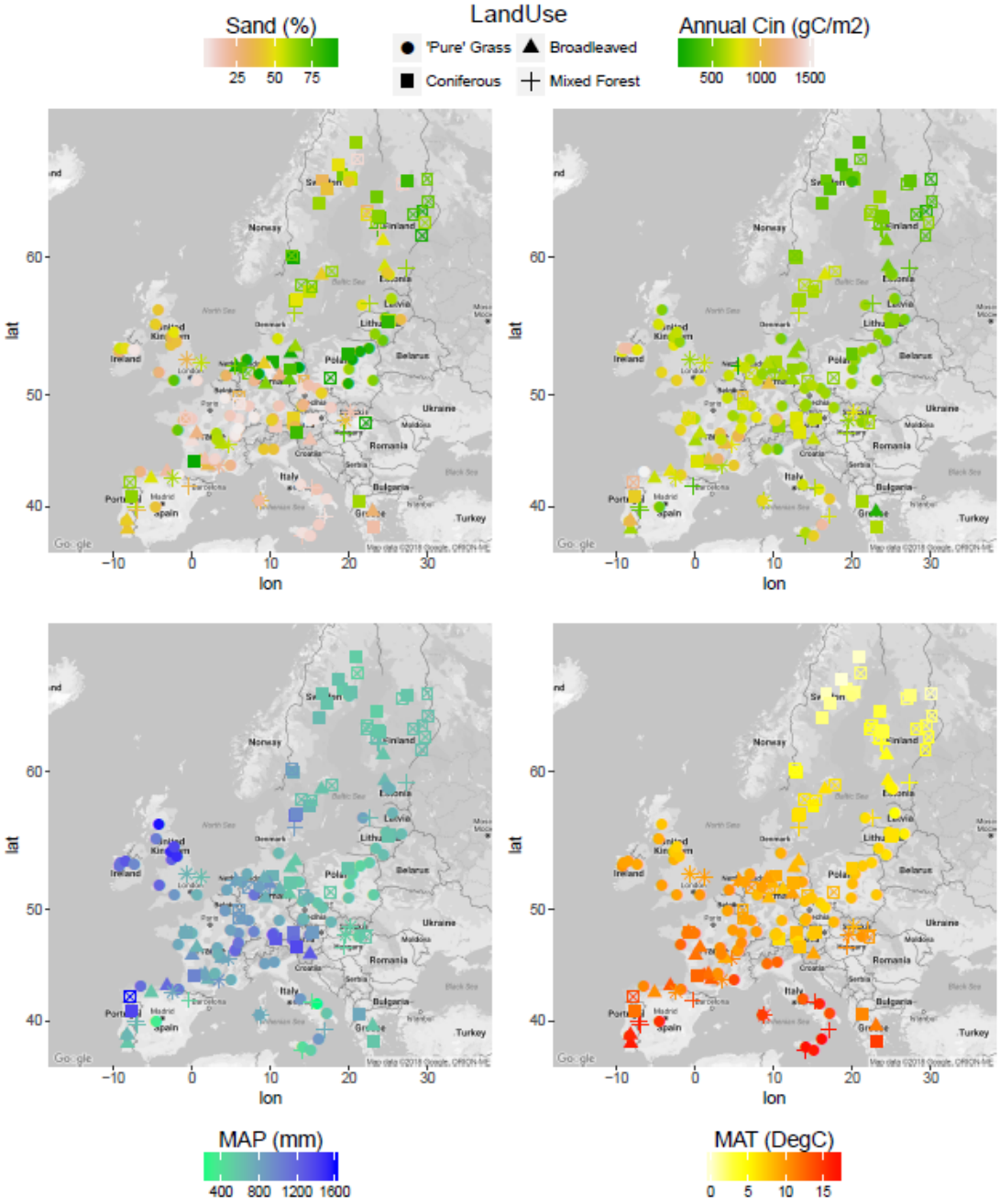


Figure S3 - Summary statistics of the site information and soil C stocks for four land use classes (Grassland, n=78; Broadleaved forest, n=25; Coniferous forest, n=27; Mixed forest, n=24) across Europe. Boxplots indicate the median, first and third quartiles with the box and maximum and minimum at the extent of the whiskers. Outliers beyond the 95% are shown by individual points. MAT = Mean Annual Temperature; MAP = Mean Annual Precipitation; NPP = Net Primary Productivity; SOC = Soil Organic Carbon; POM = Particulate Organic Matter; MAOM = Mineral-Associated Organic Matter.

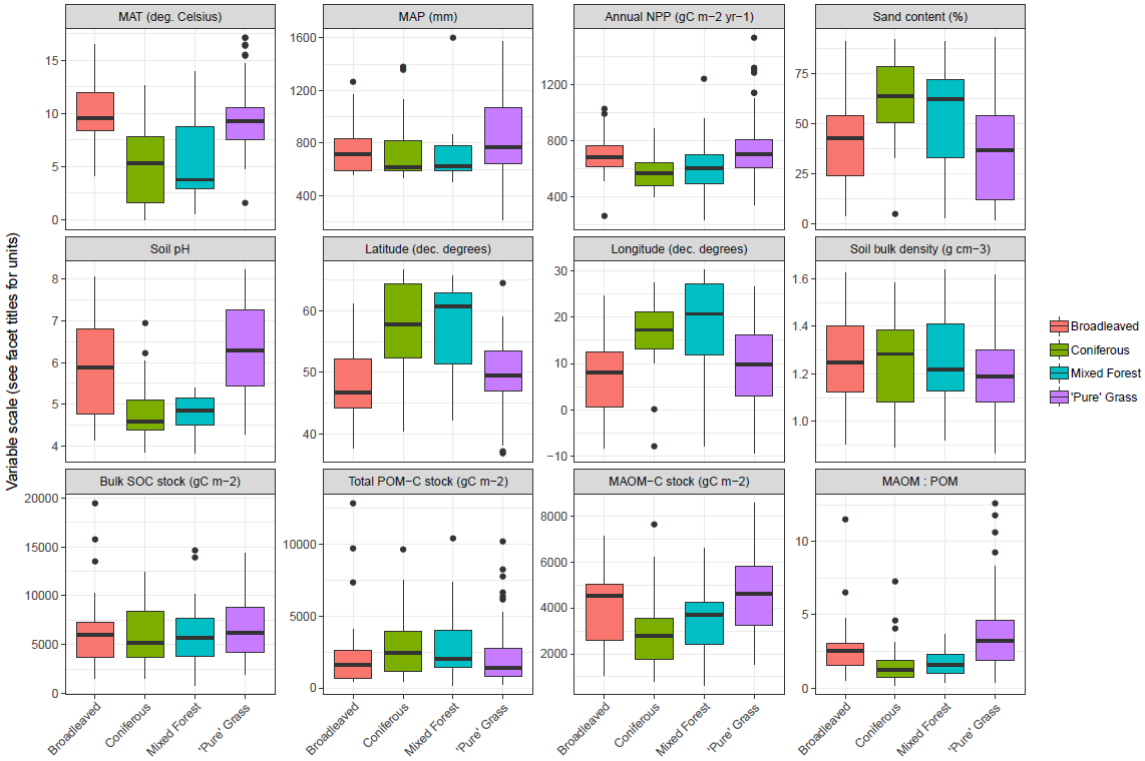


Figure S4 - One-way ANOVA results with pairwise comparisons for each measured fractionation data (bulk soil C stock, mineral-associated organic matter (MAOM) C stock, particulate organic matter (POM) C stock, and the MAOM:POM ratio) between the four land use classes (Grassland, n=78; Broadleaved forest, n=25; Coniferous forest, n=27; Mixed forest, n=24) of topsoils (0-20 cm) from 154 sites across Europe. Significant differences indicated by p-values for each pair ($p < 0.001$, red; $p < 0.01$, orange; $p < 0.05$, yellow; $p < 0.1$, green; $p > 0.1$, blue). NPP = Net Primary Productivity.

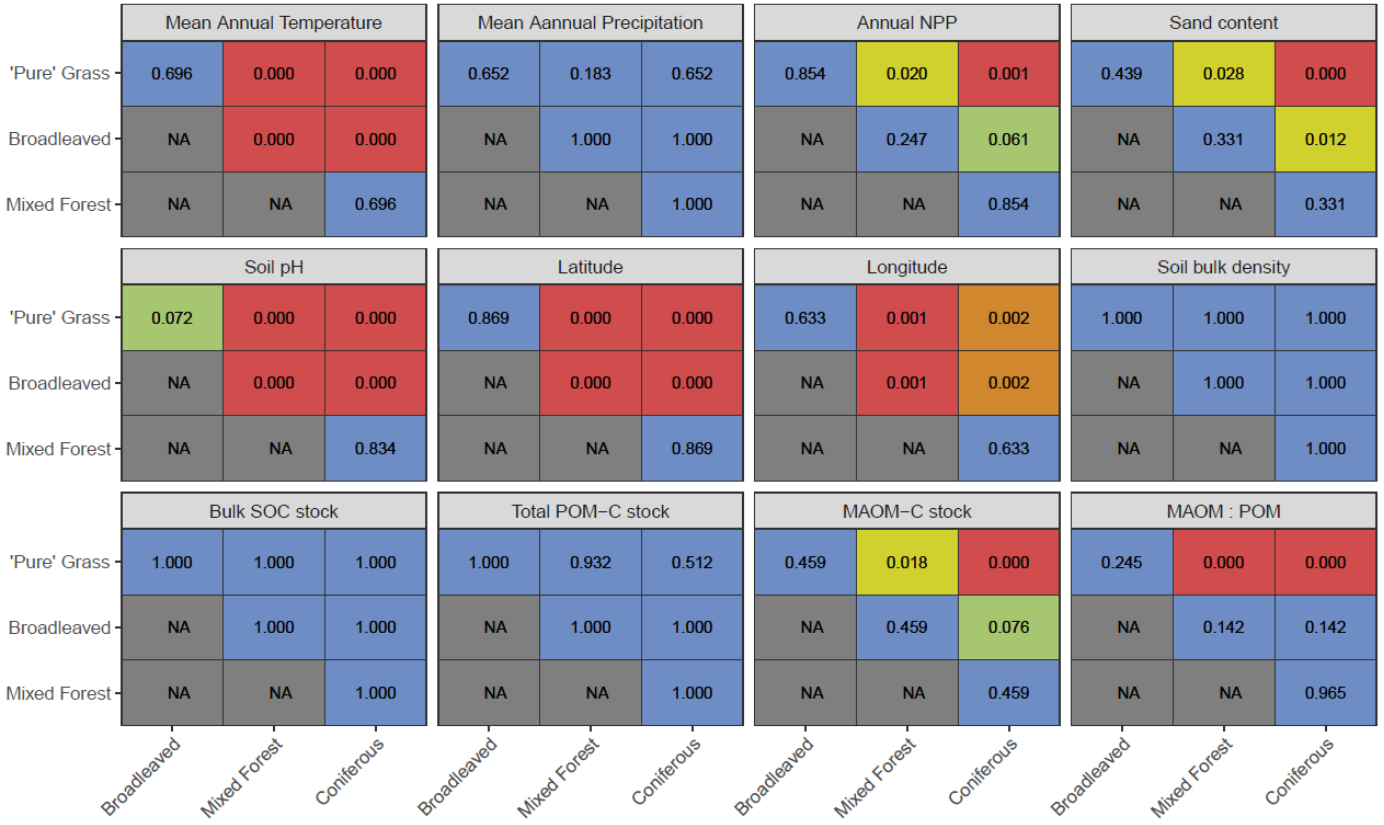


Figure S5 – Fully-colourised version of main text Figure 2. Global sensitivity analysis results showing the relative contribution of each parameter to a change in carbon stock of each pool in MEMS v1.0 (leached carbon to deeper soil layers [pool C11] is omitted for clarity). Details of each parameter and the abbreviations used can be found in Table 2. The sensitivity analysis was repeated annually for simulation times between 1 and 100 years, every 10 years after that to 400-year simulations and every 100 years after that up to a 1000-year simulation. Results are presented on a log scale in years. Parameters involved in different SOM formation processes are grouped by colour: yellows – parameters that define DOM leaching from the organic horizon to the soil layer; reds – parameters that affect microbial carbon use efficiency, purples – parameters that affect organic matter vertical transport to deeper layers, greens – maximum decay rates.

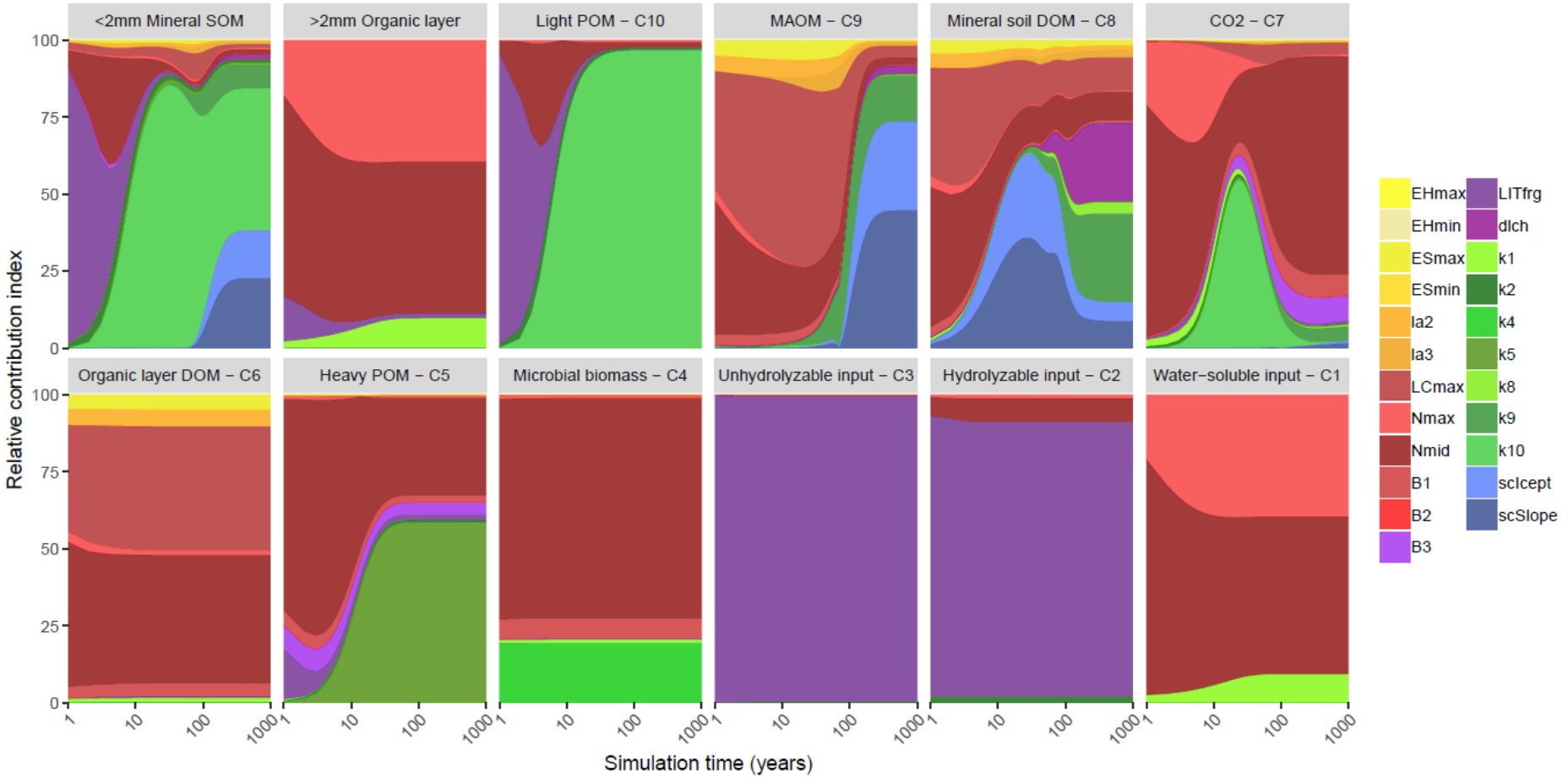


Figure S6 – Variability in model-data residuals compared with mean annual temperature for 8192 forest and grassland sites of the LUCAS dataset (Toth *et al*., 2013) simulated with the MEMS v1.0 soil organic matter model. Residuals indicate the modelled minus measured total topsoil (0-20 cm) organic carbon stock in MgC ha^{-1} for each of four land-use classes (Grassland, red; Broadleaved forest, blue; Coniferous forest, purple; Mixed forest, green). Sites are divided into high and low groups of mean annual precipitation, MAP (top *vs* bottom panels), soil texture (left *vs* right panels) and annual carbon inputs (provided by net primary productivity, NPP) (alternating panels left to right).

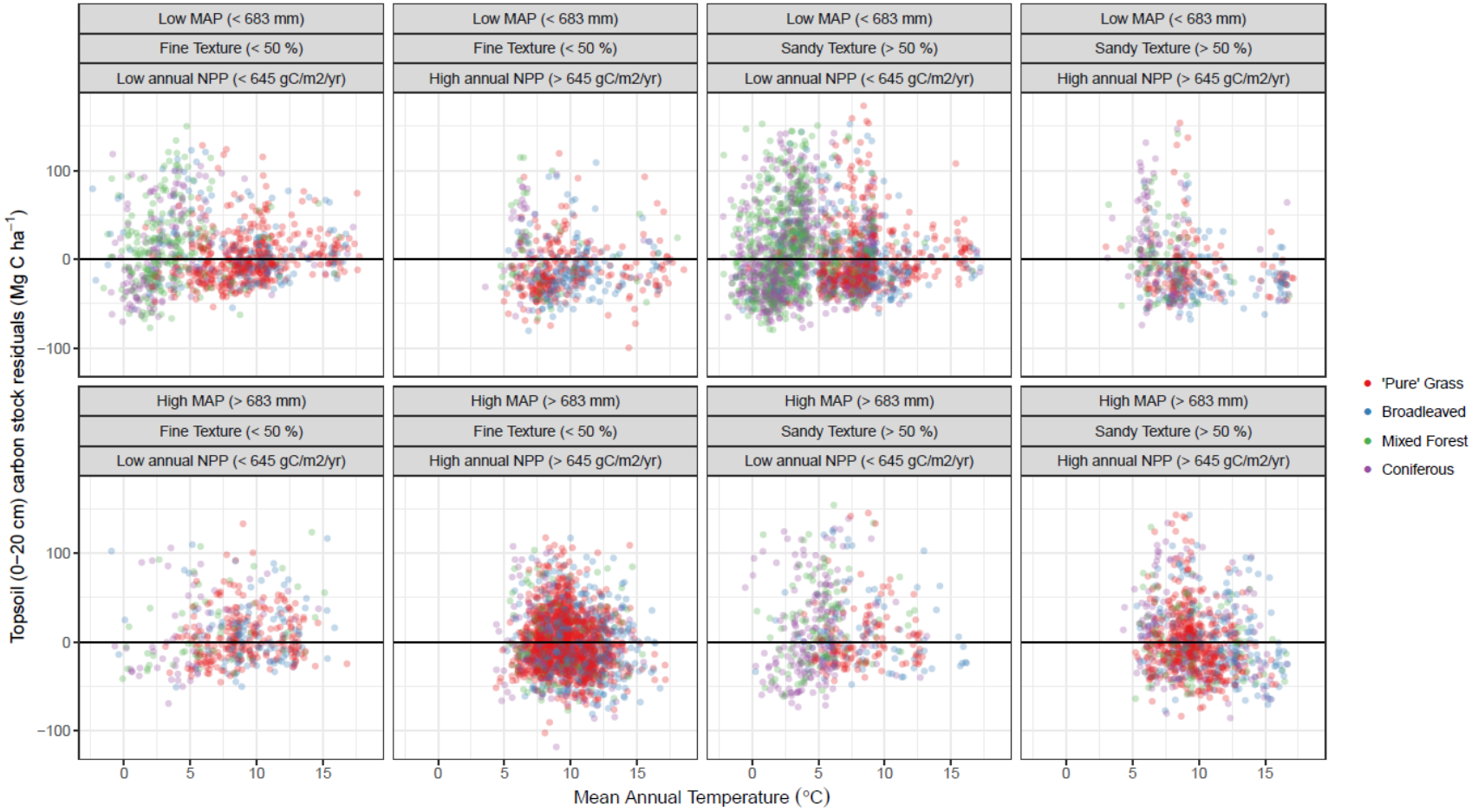
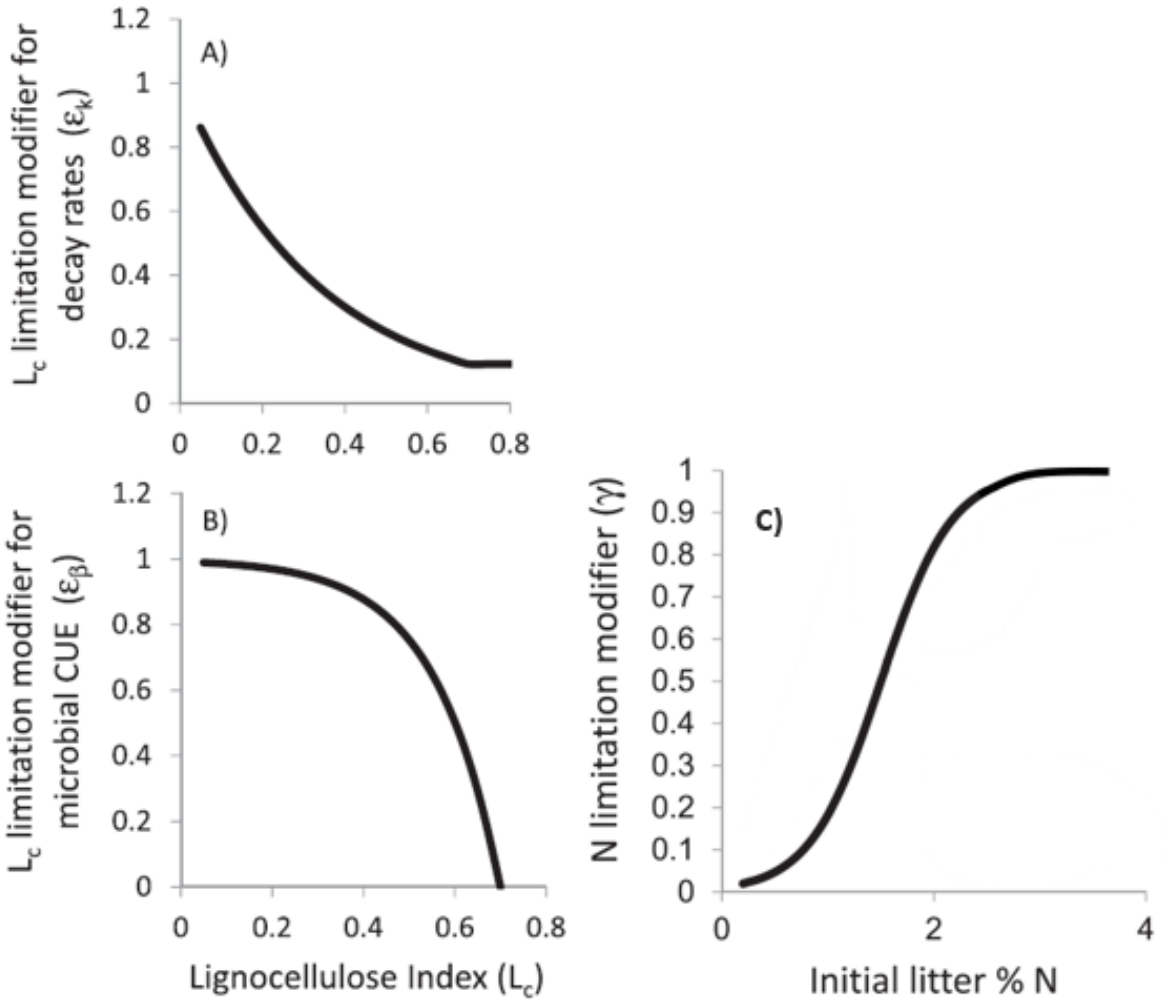


Figure S7 - Modifiers for microbial carbon use efficiency and rates of water-soluble and acid-soluble litter fractions decay by lignocellulosic index (A and B) and initial litter percent nitrogen (C). Reproduced with permission from Campbell *et al*., 2016.



SUPPLEMENTARY TABLES

Table S1 - Fractionation scheme to measure each OM pool of MEMS v1.0. Physical particle size is given sequentially from top to bottom (i.e. C9 pools are between 0.45 µm and 53 µm in size). Soil particles (< 2mm) are primary particles obtained after soil aggregates dispersion. All SOM fractions can be separated sequentially on one soil sample by first isolating the DOM through centrifugation, separating the solid subnatant into a light POM and a heavy fraction by density (at 1.8 g/cm^3) and the latter into a heavy POM and a MAOM by wet sieving (at 53µm). NDF – Neutral detergent fibre; ADF – Acid detergent fibre; HWE – Hot-water extractable.

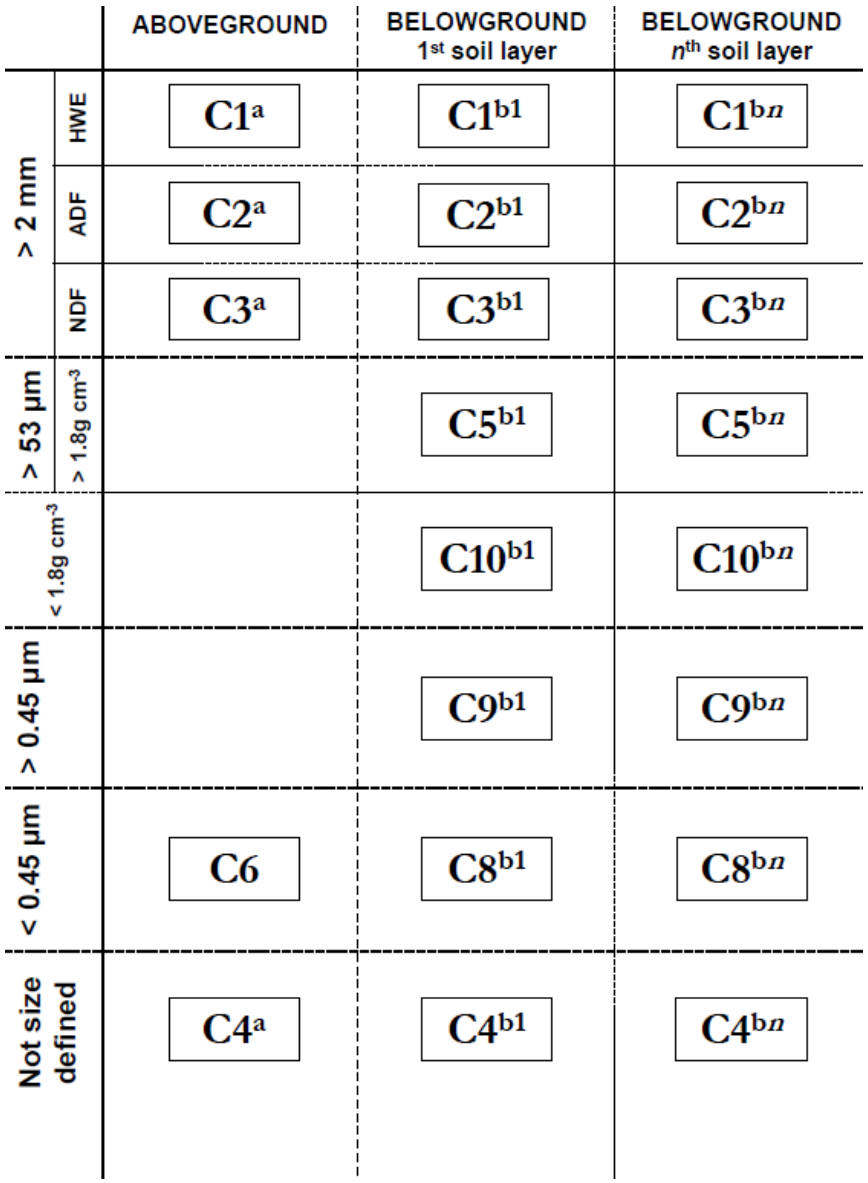


Table S2 - Optimized parameter values for the mid-point of the nitrogen modifier (*Nmid*), maximum decay rate for coarse, heavy particulate organic matter (*k5*), maximum decay rate for mineral-associated organic matter (*k9*) and maximum decay rate for light particulate organic matter (*k10*). Depending on what fraction was match (measured-modelled comparisons), different parameter values were derived. Root mean square error (RMSE) was minimised for each unique parameter set and assessed for each fraction (Mineral-Associated Organic Matter, MAOM; total Particulate Organic Matter, POM; bulk soil Soil Organic Carbon, SOC). Note that total POM refers to the composite of light and heavy POM measurements and the sum of the C5 and C10 pools). Analysis was performed on 154 forest and grassland sites from the LUCAS database – see Figure S2 and Figure S3 for more information.

Parameter	**Default (Initial optimized range)**	**Optimized for POM**	**Optimized for MAOM**	**Optimized for total SOC**
Nmid	1.750 (0.875 – 2.625)	1.617	0.923	2.448
k5	5.00^{-4} (6.0^{-5} – 1.0^{-3})	5.66^{-4}	2.37^{-4}	2.51^{-4}
k9	2.19^{-5} (1.0^{-5} – 4.0^{-5})	2.33^{-5}	2.98^{-5}	3.97^{-5}
k10	2.96^{-4} (1.0^{-4} – 1.0^{-3})	4.31^{-4}	2.93^{-4}	3.01^{-4}
RMSE between measured and modelled C stocks for 154 sites (Mg C ha^{-1})				
Total SOC	35.5	35.9	35.2	33.5
POM-C	23.4	23.5	23.1	25.5
MAOM-C	17.9	17.8	17.5	20.2

SUPPLEMENTARY REFERENCES

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., Finzi, A., Moorhead, D., Schimel, J., Torn, M. & Mayes, M. A.: The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, 137(1-2), 51-71, 2018.

Bolinder, M. A., Janzen, H. H., Gregorich, E. G., Angers, D. A., & VandenBygaart, A. J.: An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. *Agriculture, Ecosystems & Environment*, *118*(1-4), 29-42, 2007.

Braakhekke, M. C., Beer, C., Hoosbeek, M. R., Reichstein, M., Kruijt, B., Schrumpf, M., & Kabat, P.: SOMPROF: A vertically explicit soil organic matter model. *Ecological modelling,* 222(10), 1712-1730, 2011.

Campbell, E. E., Parton, W. J., Soong, J. L., Paustian, K., Hobbs, N. T., & Cotrufo, M. F.: Using litter chemistry controls on microbial processes to partition litter carbon fluxes with the litter decomposition and leaching (LIDEL) model. *Soil Biology and Biochemistry*, 100, 160-174, 2016.

Canadell, J., Jackson, R. B., Ehleringer, J. B., Mooney, H. A., Sala, O. E., & Schulze, E. D.: Maximum rooting depth of vegetation types at the global scale. *Oecologia*, 108(4), 583-595, 1996.

Harmon, M., and Domingo, J.: A User's Guide to STANDCARB Version 2.0: A Model to Simulate the Carbon Stores in Forest Stands, Dep. of For. Sci., Oreg. State Univ., Corvallis, 2001.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E., & Schulze, E. D.: A global analysis of root distributions for terrestrial biomes. *Oecologia*, 108(3), 389-411, 1996.

Kalbitz, K., Schwesig, D., Rethemeyer, J., & Matzner, E.: Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biology and Biochemistry*, 37(7), 1319-1331, 2005.

Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H., Menichetti, L.: Roots contribute more to refractory soil organic matter than aboveground crop residues, as revealed by a long-term field experiment. *Agriculture Ecosystems and Environment*, 141(1-2), 184–192, 2011.

Keiluweit, M., Bougoure, J. J., Nico, P. S., Pett-Ridge, J., Weber, P. K., & Kleber, M.: Mineral protection of soil carbon counteracted by root exudates. *Nature Climate Change*, *5*(6), 588, 2015.

Kothawala, D. N., Moore, T. R., & Hendershot, W. H.: Adsorption of dissolved organic carbon to mineral soils: A comparison of four isotherm approaches. *Geoderma*, 148(1), 43-50, 2008.

Ludwig, M., Achtenhagen, J., Miltner, A., Eckhardt, K. U., Leinweber, P., Emmerling, C., & Thiele-Bruhn, S.: Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. *Soil Biology and Biochemistry*, *81*, 311-322, 2015.

Mayes, M. A., Heal, K. R., Brandt, C. C., Phillips, J. R., & Jardine, P. M.: Relation between soil order and sorption of dissolved organic carbon in temperate subsoils. *Soil Science Society of America Journal*, 76(3), 1027-1037, 2012.

Moorhead, D. L., Lashermes, G., Sinsabaugh, R. L., & Weintraub, M. N.: Calculating co-metabolic costs of lignin decay and their impacts on carbon use efficiency. *Soil Biology and Biochemistry*, 66, 17-19, 2013.

Poeplau, C.: Estimating root: shoot ratio and soil carbon inputs in temperate grasslands with the RothC model. *Plant and soil*, 407(1-2), 293-305, 2016.

R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2018.

Sinsabaugh, R. L., Manzoni, S., Moorhead, D. L., & Richter, A.: Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling. *Ecology letters*, 16(7), 930-939, 2013.

Six, J., Conant, R. T., Paul, E. A., & Paustian, K.: Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. *Plant and soil*, 241(2), 155-176, 2002.

Sokol, N. W., Sanderman, J., & Bradford, M. A.: Pathways of mineral-associated soil organic matter formation: Integrating the role of plant carbon source, chemistry, and point of entry. *Global change biology*. https://doi.org/10.1111/gcb.14482, 2018.

Soong, J. L., Parton, W. J., Calderon, F., Campbell, E. E., & Cotrufo, M. F.: A new conceptual model on the fate and controls of fresh and pyrolized plant litter decomposition. *Biogeochemistry*, 124(1-3), 27-44, 2015.

Toth G., Jones A., Montanarella L.: LUCAS Topsoil Survey — methodology, data and results. In: JRC Technical Reports. European Union, Luxemburg, 2013.