

1 **SUPPLEMENTARY MATERIALS FOR:**

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

3

4

## 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.

$$\frac{dC1}{dt} = C1_{in}^i - (uk * C1 * k_1) \quad (1)$$

$$\frac{dC2}{dt} = C2_{in}^i - (uk * C2 * k_2) - (C2 * LIT_{frag}) \quad (2)$$

$$\frac{dC3}{dt} = C3_{in}^i - (C3 * k_3) - (C3 * LIT_{frag}) \quad (3)$$

$$\frac{dC4}{dt} = C4_{ass}^{C1} + C4_{ass}^{C2} - (C4 * k_4) \quad (4)$$

$$\frac{dC5}{dt} = C5_{gen}^{C4} + C5_{frag}^{C2} + C5_{frag}^{C3} - (C5 * k_5) \quad (5)$$

$$\frac{dC6}{dt} = C6_{in}^i + C6_{in}^{C1} + C6_{in}^{C2} + C6_{in}^{C3} + C6_{in}^{C4} - C8_{in}^{C6} \quad (6)$$

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

$$\frac{dC8}{dt} = C8_{in}^{C5} + C8_{in}^{C6} + C8_{in}^{C10} - sorption - (C8 * DOC_{lch}) - (C8 * k_8) \quad (8)$$

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

$$\frac{dC10}{dt} = C10_{frag}^{C2} + C10_{frag}^{C3} - (C10 * k_{10}) \quad (10)$$

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

## 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 ( $f_{LIG}^i$ ).

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

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

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

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

Where  ${}^L_j X_{in}^i$  is refers to the daily carbon input to pool  $X$  from external source  $i$  for layer  $L$  on day  $j$ , and  ${}^L_j CT^i$  is the total daily carbon input from external source  $i$  for layer  $L$  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).

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 carbon dioxide (CO<sub>2</sub>) emissions. The primary carbon losses to CO<sub>2</sub> 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 of by microbial biomass (C<sub>4</sub>) in the litter layer results from the balance between anabolic and catabolic processes and thus as biomass is formed there is also CO<sub>2</sub> as well as carbon in dissolved organic matter (DOM) production. Microbial assimilation is a function of nitrogen content and lignocellulosic index (Eq. 16) of the structural litter pools (C<sub>2</sub> and C<sub>3</sub>; organic matter > 2 mm) in each layer and controlled by maximum decomposition rates for C<sub>1</sub> ( $k_1$ ) and C<sub>2</sub> ( $k_2$ ) that assume first-order decay.

$${}^L_jLCI_{lit} = \frac{{}^L_jC_3}{({}^L_jC_2 + {}^L_jC_3)} \quad (16)$$

$${}^L_jC_4^{C_1}_{ass} = uB * B_1 * (1 - la_4) * uk * k_1 * {}^L_jC_1 \quad (17)$$

$${}^L_jC_4^{C_2}_{ass} = uB * B_2 * (1 - la_1) * uk * k_2 * {}^L_jC_2 \quad (18)$$

Where  ${}^L_jC_4^{C_1}_{ass}$  and  ${}^L_jC_4^{C_2}_{ass}$  refer to the fraction of the given litter pool (i.e., C<sub>1</sub> or C<sub>2</sub>) that is microbially assimilated to pool C<sub>4</sub> of layer  $L$  on day  $j$  from pool C<sub>1</sub> or C<sub>2</sub>, respectively. Note that these functions are specific to a single layer (aboveground litter in MEMS v1.0) and parameter values may differ between layers, when more are added. More information of the parameters  $uB$ ,  $uk$ ,  $B_x$ ,  $la_x$  and  $k_x$  can be found in Campbell *et al.* (2016) and Table 2 in the main manuscript, but briefly:

- ${}^L_juB$  and  ${}^L_juk$  are rate modifiers to represent the litter chemistry controls (LCI and available nitrogen) on microbial use efficiency, for layer  $L$  on day  $j$ .

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

$${}^L_j uK = \min \left( \left( \frac{1}{1 + e^{-N_{max}(N_{lit} - N_{mid})}} \right), \left( e^{-3 * {}^L_j LCI_{lit}} \right) \right) \quad (20)$$

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 S6).  $N_{lit}$  and  ${}^L_j LCI_{lit}$  are the input material nitrogen content and LCI of layer  $L$  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  ${}^L_j N_{lit}$  – as equivalent to  ${}^L_j LCI_{lit}$ ) 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 of layer  $L$  on day  $j$ .

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

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

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  ${}^L_j LCI_{lit}$  are the nitrogen content of input material and LCI of layer  $L$  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  $> 53 \mu m$  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.

$${}^L_j C5_{gen}^{C4} = B_3 * (1 - la_2) * k_4 * {}^L_j C4 \quad (23)$$

Where  ${}^L_j C5_{gen}^{C4}$  refers to the fraction of carbon that is transferred from C4 to C5 (i.e., microbial products transported belowground with structural litter fragmentation and bioturbation or advection and leaching of DOC for layer  $L$  on day  $j$ ). Belowground, this flux does not move vertically between layers but is transferred from C4 to C5 within the same soil layer. 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 ( $LIT_{frag}$ ). 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.

$${}^L_jC5^{C2}_{frag} = POM_{split} * LIT_{frag} * {}^L_jC2 \quad (24)$$

$${}^L_jC5^{C3}_{frag} = POM_{split} * LIT_{frag} * {}^L_jC3 \quad (25)$$

$${}^L_jC10^{C2}_{frag} = (1 - POM_{split}) * LIT_{frag} * {}^L_jC2 \quad (26)$$

$${}^L_jC10^{C3}_{frag} = (1 - POM_{split}) * LIT_{frag} * {}^L_jC3 \quad (27)$$

Where  ${}^L_jCX^{CY}_{frag}$  refers to the amount of carbon that is transferred from pool CY to pool CX for layer L 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;  ${}^L_jC6_{in}^i$ ), from all litter layer pools ( ${}^L_jC6_{in}^{C1}$ ,  ${}^L_jC6_{in}^{C2}$ ,  ${}^L_jC6_{in}^{C3}$ ) and from microbial biomass ( ${}^L_jC6_{in}^{C4}$ ).

$${}^L_jC6_{in}^{C1} = la_4 * uk * k_1 * {}^L_jC1 \quad (28)$$

$${}^L_jC6_{in}^{C2} = la_1 * uk * k_2 * {}^L_jC2 \quad (29)$$

$${}^L_jC6_{in}^{C3} = la_3 * k_3 * {}^L_jC3 \quad (30)$$

$${}^L_jC6_{in}^{C4} = la_2 * k_4 * {}^L_jC4 \quad (31)$$

Where  ${}^L_jC\chi_{in}^{cy}$  refers to DOM leaching from pool  $y$  to pool  $x$  of layer  $L$  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 only exists in the aboveground litter layer and therefore in the above equations  $L$  is always the aboveground layer. However, measurably, aboveground litter layer DOM is directly equivalent to the belowground soil DOM (C8). In MEMS v1.0, DOM enters the soil through the C6 pool only. However, when explicit inputs from belowground litter (e.g., roots) are simulated in future versions Eqs. 28-31 can apply for each soil layer adding 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 ( ${}^L_jC8_{in}^{C5}$ ,  ${}^L_jsorption$ ,  ${}^L_jC8_{in}^{C10}$ ) as well as from leached litter DOM (C6). Here, the *sorption* flux represents the net carbon exchange between soil DOM (C8) and MAOM (C9).

$${}^L_jC8_{in}^{C5} = la_3 * k_5 * {}^L_jC5 \quad (32)$$

$${}^L_jC8_{in}^{C6} = DOC_{frag} * {}^L_jC6 \quad (33)$$

$${}^L_jC8_{in}^{C10} = la_3 * k_{10} * {}^L_jC10 \quad (34)$$

The parameter values are defined in Table 2 in the main manuscript. As with the  $LIT_{frag}$  parameter, the  $DOC_{frag}$  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.

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

Where  ${}^L soilpH$  refers to the ‘native’ soil pH of simulated soil layer  $L$ . 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.,  $100 - sand$ )

$${}^LQ_{max} = {}^L\rho * (0.26126 * (100 - {}^Lsand) + 11.07820) * (1 - {}^Lrock) \quad (36)$$

Where  ${}^L\rho$  refers to the bulk density of soil layer  $L$  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<sup>-2</sup> layer depth<sup>-1</sup>) and therefore the equations shown here assume a 1 meter deep layer for simplification. Both the sand content ( ${}^Lsand$ ) and rock fraction ( ${}^Lrock$ ) are expressed in percent (i.e., 0-100) and specific to layer  $L$ . 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.

$${}^Lj_{sorption} = {}^LjC8 * \frac{\left( \left( \frac{({}^LK_{lm} * {}^LQ_{max} * {}^LjC8)}{1 + ({}^LjK_{lm} * {}^LjC8)} \right) - {}^LjC9 \right)}{{}^LQ_{max}} \quad (37)$$

Where  ${}^Lj_{sorption}$  is a net exchange of carbon between the soil DOM (C8) and MAOM (C9) pools of layer  $L$  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<sub>2</sub> 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<sub>2</sub> emissions which continually accumulate in the sink C7. The amount of CO<sub>2</sub> 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<sub>2</sub> emissions, as is more common in traditional SOM models (e.g., CENTURY or RothC). As with earlier equations, these below are layer- and time-specific but for simplicity are presented in a generalized form.

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

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

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

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

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

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

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

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

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  $k_3$  and  $k_5$  were best estimated in relation to the maximum decay rate of the microbially-accessible litter (C2) pool ( $k_2$ ).

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

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

Note that when  $k_2$  is a fixed value,  $k_3$  only fluctuates with changes in the LCI of the litter layer. At present, CO<sub>2</sub> emitted from soil DOM (determined by the maximum decay rate,  $k_8$ ) is associated with the values presented in Kalbitz *et al.* (2005). Also note that because the maximum decay rate of acid-insoluble litter ( $k_3$ ) is determined relative to the LCI of all litter pools in a given layer ( $L$ ) on a given day ( $j$ ) the parameter itself is also layer- and time-specific.

#### 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<sub>10</sub> 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 ( $T_{lag}$ ).

$${}_j^L T_{mod} = e^{\left( - \left( \frac{{}_j^L T_{soil}}{T_{opt} + T_{lag}} \right) \right)^{T_{shp}}} * T_{Q10}^{\frac{{}_j^L T_{soil} - T_{ref}}{T_{ref}}} \quad (48)$$

Where  ${}^L_jT_{mod}$  is the temperature multiplier applied to decomposition of pools in layer  $L$  on day  $j$ , given the soil temperature of that layer on that day ( ${}^L_jsoilT$ ). 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  $x$  of layer  $L$  on day  $j$  – e.g.  ${}^L_jT_{mod}^x$ . 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 1<sup>st</sup>). 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.

$$T_{soil} = \frac{T_{range}}{2} * \sin((2 * PI * seq) - 1.5) + MAT \quad (49)$$

Where  ${}_j^L soilT$  is the soil temperature in degrees Celsius for soil layer  $L$  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,  $PI_{seq}$  is a sequence of 365 values evenly distributed from 0 to  $\pi$  ( $\approx 3.14159$ ), 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 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.

$${}_jaCT^i = {}_jCT^i * \left( \frac{1}{RtoS^i + 1} \right) * (1 - {}_jaHARV^i) \quad (51)$$

$${}_jbCT^i = {}_jCT^i * \left( \frac{RtoS^i}{RtoS^i + 1} \right) * (1 - {}_LjbHARV^i) \quad (52)$$

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.

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

Where  ${}_j^L bCT^i$  is the belowground carbon input of material  $i$  to soil layer  $L$  on day  $j$ ,  ${}^L depth$  is the depth of soil layer  $L$  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).

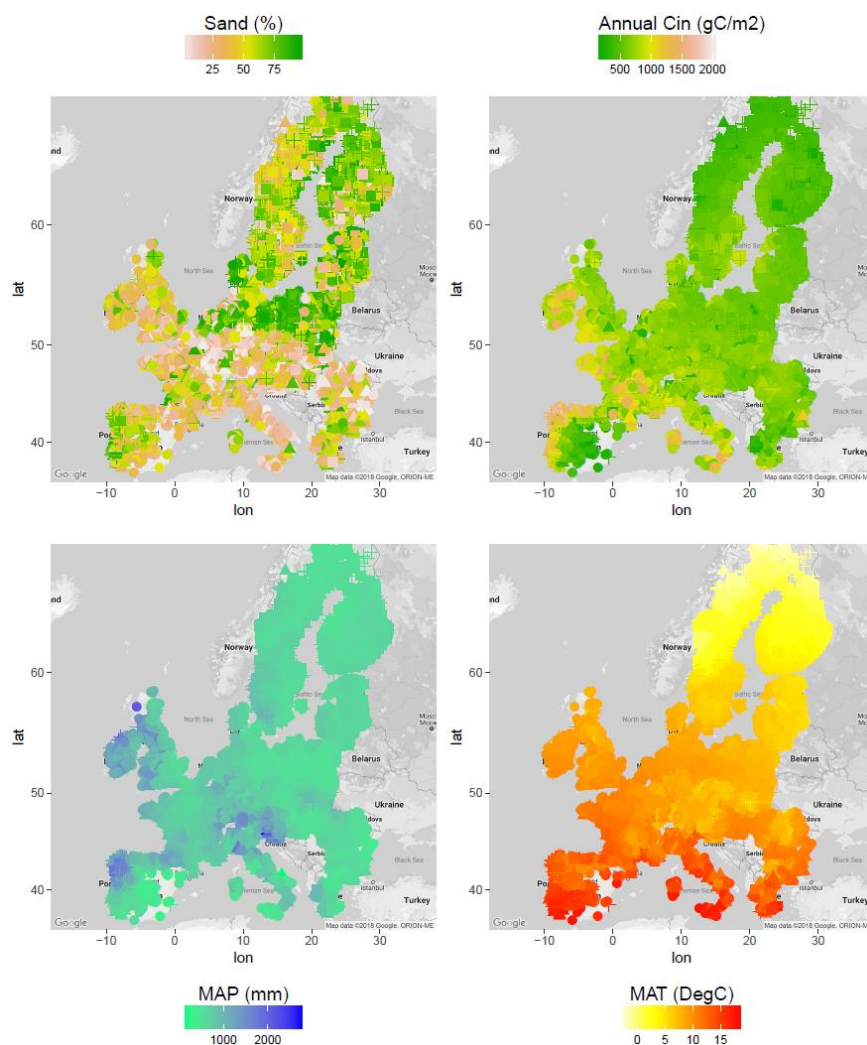
381 As with the interpolation of daily soil temperature from MAT, estimating daily values of carbon  
382 input are less precise than using real measured data. When possible, empirical data should be  
383 preferred and can be input along with daily climate data.

384

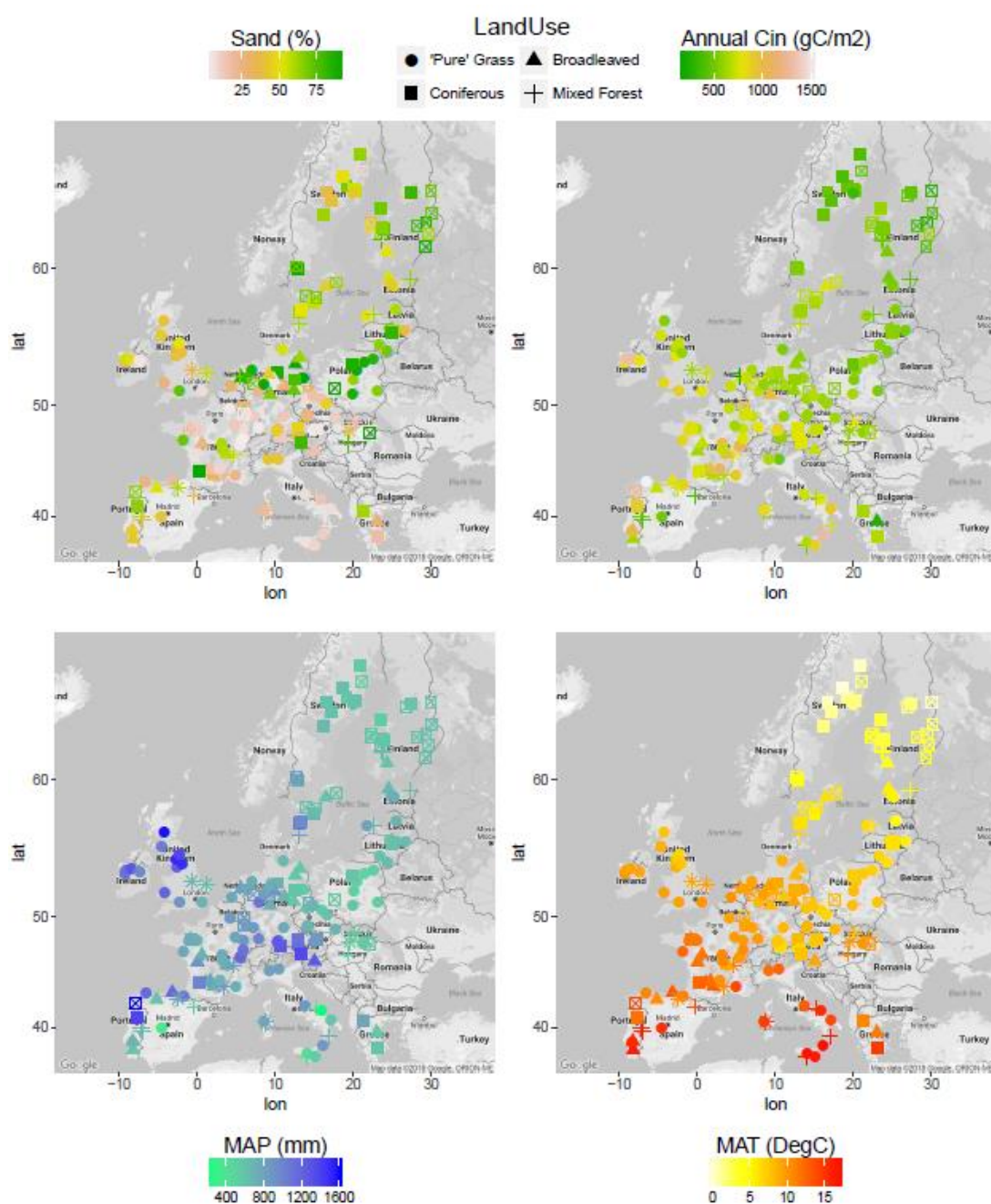
## SUPPLEMENTARY FIGURES

(see attached files for high-resolution versions)

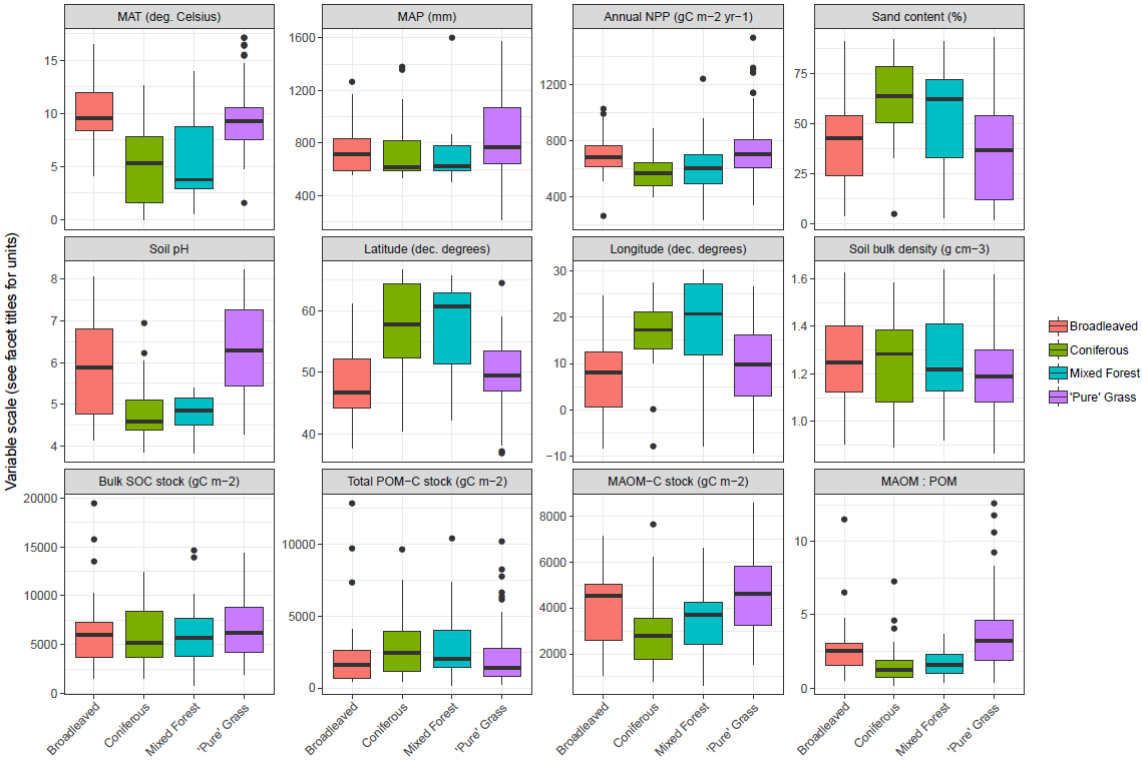
**Figure S1** – Site information of all 8192 forest and grassland sites of the LUCAS dataset (Toth *et al.*, 2013) used for validation of the MEMS v1.0 soil organic matter model. Different shapes represent different land use classes and all are overlaid over each other (grass = circles,  $n = 3487$ ;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



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



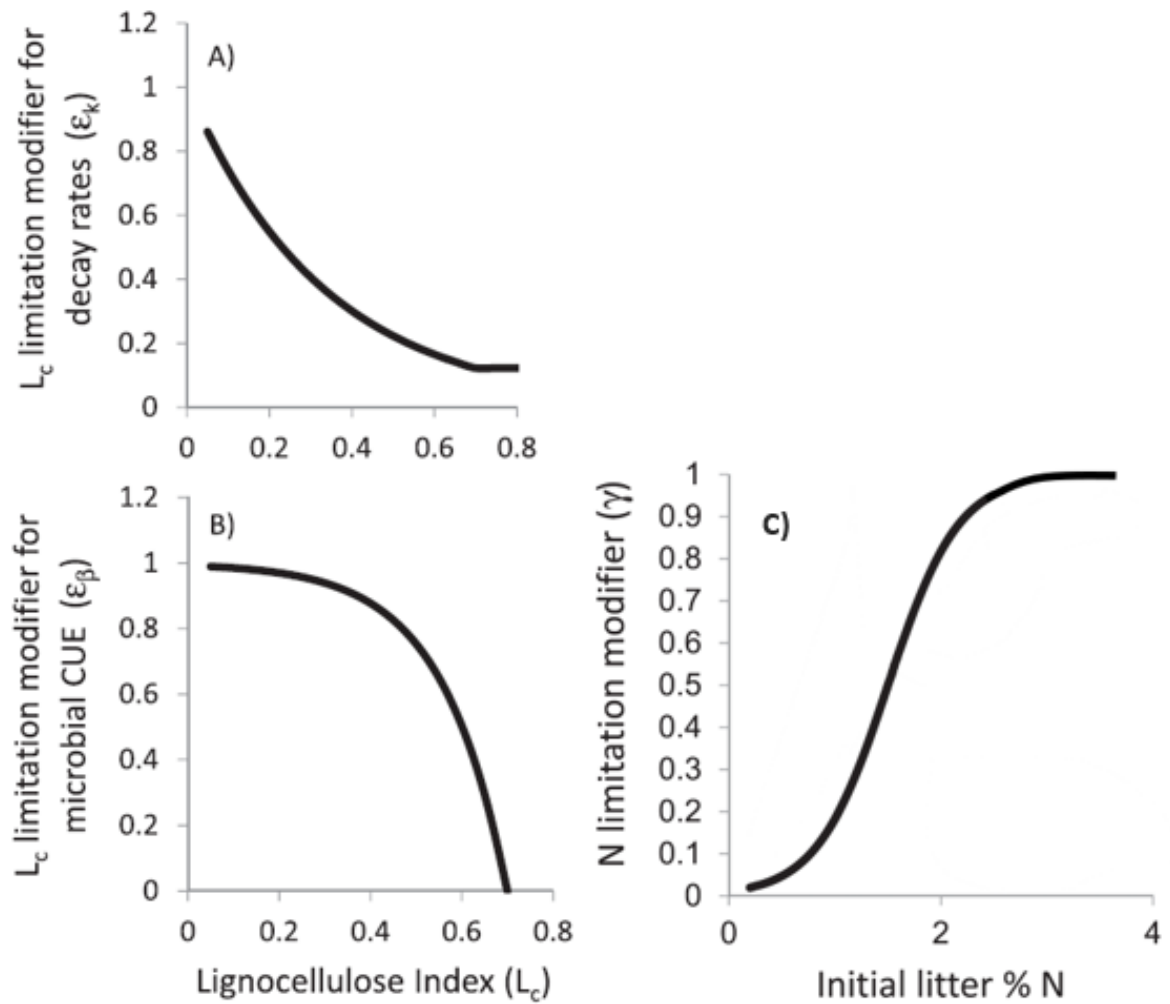
410  
 411

412 **Figure S4** - One-way ANOVA results with pairwise comparisons for each measured fractionation data (bulk soil C stock, mineral-associated  
413 organic matter (MAOM) C stock, particulate organic matter (POM) C stock, and the MAOM:POM ratio) between the four land use classes  
414 (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.  
415 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  
416 = Net Primary Productivity.

	Mean Annual Temperature			Mean Annual Precipitation			Annual NPP			Sand content		
'Pure' Grass	0.696	0.000	0.000	0.652	0.183	0.652	0.854	0.020	0.001	0.439	0.028	0.000
Broadleaved	NA	0.000	0.000	NA	1.000	1.000	NA	0.247	0.061	NA	0.331	0.012
Mixed Forest	NA	NA	0.696	NA	NA	1.000	NA	NA	0.854	NA	NA	0.331
	Soil pH			Latitude			Longitude			Soil bulk density		
'Pure' Grass	0.072	0.000	0.000	0.869	0.000	0.000	0.633	0.001	0.002	1.000	1.000	1.000
Broadleaved	NA	0.000	0.000	NA	0.000	0.000	NA	0.001	0.002	NA	1.000	1.000
Mixed Forest	NA	NA	0.834	NA	NA	0.869	NA	NA	0.633	NA	NA	1.000
	Bulk SOC stock			Total POM-C stock			MAOM-C stock			MAOM : POM		
'Pure' Grass	1.000	1.000	1.000	1.000	0.932	0.512	0.459	0.018	0.000	0.245	0.000	0.000
Broadleaved	NA	1.000	1.000	NA	1.000	1.000	NA	0.459	0.076	NA	0.142	0.142
Mixed Forest	NA	NA	1.000	NA	NA	1.000	NA	NA	0.459	NA	NA	0.965
	Broadleaved	Mixed Forest	Coniferous	Broadleaved	Mixed Forest	Coniferous	Broadleaved	Mixed Forest	Coniferous	Broadleaved	Mixed Forest	Coniferous

417

**Figure S5** - 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  $\mu\text{m}$  and 53  $\mu\text{m}$  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  $\text{g}/\text{cm}^3$ ) and the latter into a heavy POM and a MAOM by wet sieving (at 53 $\mu\text{m}$ ).  
NDF – Neutral detergent fibre; ADF – Acid detergent fibre; HWE – Hot-water extractable.

		ABOVEGROUND	BELOWGROUND 1 <sup>st</sup> soil layer	BELOWGROUND $n^{\text{th}}$ soil layer
> 2 mm	HWE	C1 <sup>a</sup>	C1 <sup>b1</sup>	C1 <sup>bn</sup>
	ADF	C2 <sup>a</sup>	C2 <sup>b1</sup>	C2 <sup>bn</sup>
	NDF	C3 <sup>a</sup>	C3 <sup>b1</sup>	C3 <sup>bn</sup>
> 53 $\mu\text{m}$	> 1.8g $\text{cm}^{-3}$		C5 <sup>b1</sup>	C5 <sup>bn</sup>
	< 1.8g $\text{cm}^{-3}$		C10 <sup>b1</sup>	C10 <sup>bn</sup>
> 0.45 $\mu\text{m}$			C9 <sup>b1</sup>	C9 <sup>bn</sup>
< 0.45 $\mu\text{m}$		C6	C8 <sup>b1</sup>	C8 <sup>bn</sup>
Not size defined		C4 <sup>a</sup>	C4 <sup>b1</sup>	C4 <sup>bn</sup>

**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
<i>Nmid</i>	1.750 (0.875 – 2.625)	1.603	0.912	2.450
<i>k5</i>	$5.00^{-4}$ ( $6.0^{-5}$ – $1.0^{-3}$ )	$5.71^{-4}$	$2.36^{-4}$	$2.53^{-4}$
<i>k9</i>	$2.19^{-5}$ ( $1.0^{-5}$ – $4.0^{-5}$ )	$2.37^{-5}$	$2.97^{-5}$	$3.97^{-5}$
<i>k10</i>	$2.96^{-4}$ ( $1.0^{-4}$ – $1.0^{-3}$ )	$4.30^{-4}$	$2.94^{-4}$	$3.02^{-4}$
<b>RMSE between measured and modelled C stocks for 154 sites (Mg C ha<sup>-1</sup>)</b>				
Total SOC	35.5	35.7	35.1	33.7
POM-C	23.4	23.4	23.1	25.3
MAOM-C	17.9	17.7	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. (2018). The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry*, 137(1-2), 51-71.
- Bolinder, M. A., Janzen, H. H., Gregorich, E. G., Angers, D. A., & VandenBygaart, A. J. (2007). 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.
- Braakhekke, M. C., Beer, C., Hoosbeek, M. R., Reichstein, M., Kruijt, B., Schrumpf, M., & Kabat, P. (2011). SOMPROF: A vertically explicit soil organic matter model. *Ecological modelling*, 222(10), 1712-1730.
- Campbell, E. E., Parton, W. J., Soong, J. L., Paustian, K., Hobbs, N. T., & Cotrufo, M. F. (2016). 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.
- Canadell, J., Jackson, R. B., Ehleringer, J. B., Mooney, H. A., Sala, O. E., & Schulze, E. D. (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia*, 108(4), 583-595.
- Harmon, M., and J. Domingo (2001), 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.
- Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E., & Schulze, E. D. (1996). A global analysis of root distributions for terrestrial biomes. *Oecologia*, 108(3), 389-411.
- Kalbitz, K., Schwesig, D., Rethemeyer, J., & Matzner, E. (2005). Stabilization of dissolved organic matter by sorption to the mineral soil. *Soil Biology and Biochemistry*, 37(7), 1319-1331.
- Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H., Menichetti, L. (2011) 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.
- Keiluweit, M., Bougoure, J. J., Nico, P. S., Pett-Ridge, J., Weber, P. K., & Kleber, M. (2015). Mineral protection of soil carbon counteracted by root exudates. *Nature Climate Change*, 5(6), 588.
- Kothawala, D. N., Moore, T. R., & Hendershot, W. H. (2008). Adsorption of dissolved organic carbon to mineral soils: A comparison of four isotherm approaches. *Geoderma*, 148(1), 43-50.
- Ludwig, M., Achtenhagen, J., Miltner, A., Eckhardt, K. U., Leinweber, P., Emmerling, C., & Thiele-Bruhn, S. (2015). Microbial contribution to SOM quantity and quality in density fractions of temperate arable soils. *Soil Biology and Biochemistry*, 81, 311-322.
- Mayes, M. A., Heal, K. R., Brandt, C. C., Phillips, J. R., & Jardine, P. M. (2012). Relation between soil order and sorption of dissolved organic carbon in temperate subsoils. *Soil Science Society of America Journal*, 76(3), 1027-1037.
- Moorhead, D. L., Lashermes, G., Sinsabaugh, R. L., & Weintraub, M. N. (2013). Calculating co-metabolic costs of lignin decay and their impacts on carbon use efficiency. *Soil Biology and Biochemistry*, 66, 17-19.
- Poeplau, C. (2016). Estimating root: shoot ratio and soil carbon inputs in temperate grasslands with the RothC model. *Plant and soil*, 407(1-2), 293-305.

- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Sinsabaugh, R. L., Manzoni, S., Moorhead, D. L., & Richter, A. (2013). Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling. *Ecology letters*, 16(7), 930-939.
- Six, J., Conant, R. T., Paul, E. A., & Paustian, K. (2002). Stabilization mechanisms of soil organic matter: implications for C-saturation of soils. *Plant and soil*, 241(2), 155-176.
- Soong, J. L., Parton, W. J., Calderon, F., Campbell, E. E., & Cotrufo, M. F. (2015). A new conceptual model on the fate and controls of fresh and pyrolyzed plant litter decomposition. *Biogeochemistry*, 124(1-3), 27-44.
- Toth G., Jones A., Montanarella L. (2013) LUCAS Topsoil Survey — methodology, data and results. In: JRC Technical Reports. European Union, Luxemburg.