KEYLINK: towards a more integrative soil representation for inclusion in ecosystem scale models. II. Model description, implementation and testing

Omar Flores [Corresp., Equal first author, 1, 2], **Gaby Deckmyn** [Equal first author, 2], **Jorge Curiel Yuste** [3, 4], **Mathieu Javaux** [5, 6], **Alexei Uvarov** [7], **Sietse van der Linde** [8], **Bruno De Vos** [9], **Harry Vereecken** [6], **Juan Jiménez** [10], **Olga Vinduskova** [2], **Andrea Schnepf** [6]

[1] Biogeography and Global Change, National Museum of Natural Sciences, Consejo Superior de Investigaciones Científicas, Madrid, Spain

[2] PLECO, Department of Biology, Universiteit Antwerpen, Antwerp, Belgium

[3] BC3 - Basque Centre for Climate Change, Leioa, Spain

[4] IKERBASQUE - Basque Foundation for Science, Bilbao, Spain

[5] Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

[6] Agrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany

[7] Laboratory of Soil Zoology, A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia

[8] Forest Research, Farnham, United Kingdom

[9] Department of Environment and Climate, Research Institute for Nature and Forest, Brussels, Belgium

[10] Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas, Jaca, Spain

Corresponding Author: Omar Flores
Email address: omar.flores@uantwerpen.be

New knowledge on soil structure highlights its importance for hydrology and soil organic matter (SOM) stabilization, which however remains neglected in many wide used models. We present here a new model, KEYLINK, in which soil structure is integrated with the existing concepts on SOM pools, and elements from food web models, i.e. those from direct trophic interactions among soil organisms. KEYLINK is, therefore, an attempt to integrate soil functional diversity and food webs in predictions of soil carbon (C) and soil water balances. We present a selection of equations that can be used for most models as well as basic parameter intervals for, e.g., key pools, functional groups' biomasses and growth rates. Parameter distributions can be determined with Bayesian calibration, and here an example is presented for food web growth rate parameters for a pine forest in Belgium. We show how these added equations can improve the functioning of the model in describing known phenomena. For this, five test cases are given as simulation examples: changing the input litter quality (recalcitrance and carbon to nitrogen ratio), excluding predators, increasing pH and changing initial soil porosity. These results overall show how KEYLINK is able to simulate the known effects of these parameters and can simulate the linked effects of biopore formation, hydrology and aggregation on soil functioning. Furthermore, the results show an important trophic cascade effect of predation on the complete C cycle with repercussions on the soil structure as ecosystem engineers are

predated, and on SOM turnover when predation on fungivore and bacterivore populations are reduced. In summary, KEYLINK shows how soil functional diversity and trophic organization and their role in C and water cycling in soils should be considered in order to improve our predictions on C sequestration and C emissions from soils.

KEYLINK: towards a more integrative soil representation for inclusion in ecosystem scale models. II. Model description, implementation and testing.

Omar Flores[Corresp., Equal first author, 1, 2*], Gaby Deckmyn[Equal first author, 2], Jorge Curiel Yuste[3, 4], Mathieu Javaux[5,6], Alexei Uvarov[7], Sietse van der Linde[8], Bruno De Vos[9], Harry Vereecken[6], Juan Jiménez[10], Olga Vinduskova[2], Andrea Schnepf[6]

[1]Department of Biogeography and Global Change, National Museum of Natural Sciences - Spanish National Research Council (MNCN-CSIC), Madrid, Spain

[2]Centre of Excellence PLECO, Department of Biology, Universiteit Antwerpen, Antwerp, Belgium

[3]BC3-Basque Centre for Climate Change, Leioa, Spain

[4] IKERBASQUE - Basque Foundation for Science, Bilbao, Spain

[5]Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

[6]Agrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany

[7]Laboratory of Soil Zoology, A.N. Severtsov Institute of Ecology and Evolution, Russian Academy of Sciences, Moscow, Russia

[8]Forest Research, Farnham, United Kingdom

[9]Department of Environment and Climate, Research Institute for Nature and Forest, Brussels, Belgium

[10]ARAID, Pyrenean Institute of Ecology - Spanish National Research Council (IPE-CSIC), Jaca, Spain

Corresponding Author:

Omar Flores[1, 2]

Universiteitsplein 1, Wilrijk, 2610, Belgium

Email address: dr.flores.omar@gmail.com

Abstract

New knowledge on soil structure highlights its importance for hydrology and soil organic matter (SOM) stabilization, which however remains neglected in many wide used models. We present here a new model, KEYLINK, in which soil structure is integrated with the existing concepts on

SOM pools, and elements from food web models, i.e. those from direct trophic interactions among soil organisms. KEYLINK is, therefore, an attempt to integrate soil functional diversity and food webs in predictions of soil carbon (C) and soil water balances. We present a selection of equations that can be used for most models as well as basic parameter intervals for, e.g., key pools, functional groups' biomasses and growth rates. Parameter distributions can be determined with Bayesian calibration, and here an example is presented for food web growth rate parameters for a pine forest in Belgium. We show how these added equations can improve the functioning of the model in describing known phenomena. For this, five test cases are given as simulation examples: changing the input litter quality (recalcitrance and carbon to nitrogen ratio), excluding predators, increasing pH and changing initial soil porosity. These results overall show how KEYLINK is able to simulate the known effects of these parameters and can simulate the linked effects of biopore formation, hydrology and aggregation on soil functioning. Furthermore, the results show an important trophic cascade effect of predation on the complete C cycle with repercussions on the soil structure as ecosystem engineers are predated, and on SOM turnover when predation on fungivore and bacterivore populations are reduced. In summary, KEYLINK shows how soil functional diversity and trophic organization and their role in C and water cycling in soils should be considered in order to improve our predictions on C sequestration and C emissions from soils.

Introduction

Soil models used in ecosystem-scale modelling need to be relatively simple and fast at performing calculations. Nonetheless, carbon (C) and nutrient turnover and hydrology are extremely important for determining ecosystem productivity and C sequestration in the ecosystem. The most widely used soil models (Century, RothC) emphasize the C flow from easily degradable to stable organic compounds using first-order kinetics to describe their decay rates (Campbell and Paustian, 2015). The relevance of chemical recalcitrance, used in those models, is accepted in the early stages of litter decomposition, but that approach has been questioned on the long term soil organic matter (SOM) stabilization (Schmidt *et al*., 2011), highlighting the relevance of other processes in the physical protection of SOM within soil matrix (Deckmyn *et al*., 2020). This has led to the development of models including an explicit representation of structural effects on SOM (Kuka *et al*., 2007). Furthermore, recent studies have shown that microbial products from the transformation of plant litter are the largest contributors to stable SOM (Mambelli *et al*., 2011; Cotrufo *et al*., 2013).

The insights concerning the role of the microbial biomass in C turnover has been introduced in models such as MIcrobial-MIneral Carbon Stabilization (MIMICS) model (Wieder *et al*., 2014; Wieder *et al*., 2015) and Litter Decomposition and Leaching (LIDEL) model (Campbell *et al*., 2016). However, soil fauna and especially ecosystem engineers, i.e. organisms that create, modify or maintain habitats by changing the physical structure of the ecosystem (Jones *et al*., 1994), have also been shown to play a key role in determining C and nutrient turnover and hydrology of soils through their impact on aggregation, pore formation and bioturbation as well as their direct contribution to litter and SOM turnover (Filser *et al*., 2016; Lavelle *et al*., 2016). Several authors have highlighted the need to include soil fauna contributions to SOM dynamics into soil modelling (see review by Vereecken *et al*., 2016). This information has been used in detailed and small-scale soil models (Chertov *et al*., 2017a; Geisen *et al*., 2019), but is not

incorporated into larger-scale ecosystem models. The main difficulty is the lack of data concerning the soil, either physical, chemical or biological, and the different methods used, making parameterization of any model unsure. The goal of the KEYLINK model is to consider the soil including the main mechanisms concerning the effects of soil biota on litter and SOM transformations and hydrology through structural modifications, without increasing the number of parameters beyond what is currently available on most well-measured ecosystems (Deckmyn *et al*., 2020). We show how this model has been parameterized for a forest stand where soil fauna was never studied in detail, but many other soil and stand characteristics are well established.

The core model concept is the strong link between soil biota, soil structure and turnover (**Fig. 1**). The decay of fresh litter is dependent on the recalcitrance and carbon to nitrogen (N) ratio (C:N) of the litter, although different soil biota groups have specific sensibilities to recalcitrance and C:N ratio. For SOM, the turnover depends on the accessibility, linked to the pore size distribution, the aeration and H_2O in the pores and the aggregation (based on the model by Kuka, Franko and Rühlmann, 2007). Both SOM and litter turnover depend on temperature and humidity. Soil fauna, specifically ecosystem engineers, directly affect pore distribution besides an important effect on bioturbation. Pore distribution affects hydrology which again affects all soil processes.

The scientific background for the model is fully described in Deckmyn *et al*. (2020). Here, the related processes are formulated mathematically. Finally, we show how the model can simulate several known mechanisms of soil faunal effects such as changes in litter recalcitrance affecting fungal/bacterial ratio, changes in pH affecting earthworm populations, effects of ecosystem engineers on bioturbation and hydrology, and importance of microbivores and predators in the soil fauna food web.

Methodology

The KEYLINK model has been conceptually designed integrating the structure of the soil by its porosity, the hydrology and the C cycle through the soil food web. Those key parts of the soil interact (**Fig. 1**) determining the rates of SOM stabilization and CO_2 emissions from soil. The functions developed to represent and simulate those processes are presented here.

Structural effects

Pore size distribution determines accessibility for trophic interactions of soil fauna and soil microorganisms (**Fig. 2**), both by size and by aeration and H_2O; soil fauna changes pore size distribution and produces cracks and fissures in the soil. In the model, pore size distribution is divided into the following five categories:

- **Inaccessible pores** (< 0.1 μm in diameter)**:** pores around inaccessible C (within the micro-aggregate, organo-clay interaction). Water is held here but is not available to plants (measured from wilting point). The volume of inaccessible pores is related to the clay content and type.
- **Bacterial pores** (0.1 – 2 μm)**:** the pores within macro-aggregates and pores in loam, accessible only to bacteria. Engineer saprotrophs (e.g. earthworms) can also use SOM in these pores (and in the following pore categories, all except inaccessible pores) because they eat directly all soil.

- **Micropores** (2 – 30 µm): pores not accessible to macrofauna, mesofauna and most predators, but accessible to microfauna bacterivores and fungivores, fungi, mycorrhiza and bacteria. Water is held at field capacity but available to plants. In sandy soil and within macro-aggregates (> 250 µm), pores fall in this category.
- **Mesopores** (30 µm – 1.5 mm): pores where most soil fauna can penetrate (not macrofauna) between large macro-aggregates (>1 mm) or formed by fine roots. Mesopore volume can be determined in the field from drained water capacity (but this includes macropores). These pores are well aerated also at field capacity, but can dry out below field capacity.
- **Macropores** (> 1.5 mm): cracks or biopores formed by ecosystem engineers. They are of vital importance for soil hydrology as preferential flow through these pores has a major impact on infiltration rate. These are the first pores to have O_2 when water level is above field capacity, but dry out quickly below field capacity.

The initial values of soil porosity in the model simulations can be calculated from measured soil water retention curves, or even using models such as Saxton *et al*. (1986) that yield field capacity, porosity and wilting point from the C, clay and sand contents, or using measured bulk density (D_b). Following Malamoud *et al*. (2009), the percentage of total porosity ($P_{\%}$) can be computed from D_b and soil particle density (D_S) as shown in equation 2. D_S can be measured or is calculated from D_m = soil mineral particle density (2.65 g cm^{-3}) and D_{SOM} = organic particle density (1.35 g cm^{-3}) as:

$$D_s = \frac{100}{\frac{\%\,SOM}{D_{SOM}} + \frac{100 - \%\,SOM}{D_m}} \quad (1)$$

$$P_{\%} = \frac{D_s - D_b}{D_s} 100 \quad (2)$$

Water flow

We advise using KEYLINK model in combination with a detailed water model including preferential flow through macropores as well as good representation for matrix flow (s.a. Richards' equation). However, we show in this paper how it can be used with a simpler representation of water flow but still allowing the important dynamic interactions between pore sizes and hydrology that are fundamental to the model. A spilling bucket approach is used at a daily time-step, where water drains from a layer into the underlying layer when its water content is above field capacity in the soil matrix. However, in contrast to conventional spilling bucket models, we allow water to flow faster through macropores (before the soil matrix is saturated). Net precipitation (P_{net}) is calculated as:

$$P_{net} = P - E \quad (3)$$

where P is precipitation (mm day^{-1}) and E is evapotranspiration (mm day^{-1}) from measured or modelled data (vegetation model). Infiltration (I) is assumed to be equal to the part of precipitation entering the soil. Infiltration and runoff (P_{runoff}, mm day^{-1}) must equal P_{net} .

$$I + P_{runoff} = P_{net} \quad (4)$$

Infiltration is composed of water entering the soil matrix, water filling the macropores and water draining from macropores. Water that enters macropores remains in the macropore domain or enters the layers below. The fraction of infiltration entering macropores depends on the surface area of the macropores (SA_{macro}), assumed cylindrical. Assume measured or derived maximal infiltration rate (I_{maxMat}, mm day^{-1}) of the soil matrix. Maximal infiltration rate through macropores (I_{maxPor}, mm day^{-1}) is calculated from the volume of the pores (PV_{macro}), assumed not limiting at daily scale, plus infiltration capacity of the layer (n+1) in which the macropores end.

$$I_{maxPor} = PV_{macro} + I_{maxMat(n+1)} \quad (5)$$

If $P_{net} > (I_{maxPor} + I_{maxMat})$ runoff is calculated as:

$$P_{runoff} = P_{net} - (I_{maxPor} + I_{maxMat}) \quad (6)$$

after which calculations continue using P_{net} - P_{runoff} as net precipitation.

If $I_{maxMat} < P_{net} < (I_{maxPor} + I_{maxMat})$ the soil matrix is filled at a rate equal to the maximum infiltration rate, all other water is lost either through the macropores to the next layer or by filling macropores. If $I_{maxMat} > P_{net}$ the soil matrix is filled with water, traditional spilling bucket, but an equivalent volume is lost through macropores to the bottom layer depending on the surface area of the macropores. The total soil water volume of soil layer n, SW_n, is then limited by the total pore volume of the layer and the water already filling the pores, and is calculated as:

$$SW_n = SW_n + \min(PV_n - SW_n, I_{maxmat}(1 - SA_{macro}), Pnet(1 - SA_{macro})) \quad (7)$$

For drainage (D) to the bottom layer, the spilling bucket approach is used plus a portion of water that goes straight through the macropores, calculated from the surface area of the macropores.

$$D_n = P_{net} SA_{macro} + P_{net} -$$

$$\min(PV_n - SW_n, I_{maxmat}(1 - SA_{macro}), P_{net}(1 - SA_{macro})) \quad (8)$$

For each pore size class the fraction water filled is calculated from the water content: so always one pore size is partially saturated and all others are either saturated or dry within one layer.

C flow

The KEYLINK model combines soil organic matter modelling with soil food web modelling. The model conceptualized in **Figure 2** has 13 carbon pools (**Table S1.2** in **Supplemental File S1**), visualised by boxes. Above and belowground litter is assumed to be provided from an external source (*tree shoot* in **Figure 2**) not covered by this model. It could be given through experimental data or an external model, e.g., a tree growth model that delivers the input of litter into the litter pool. All simulations presented here were made with constant C inputs (**Table S2.6** in **Supplemental File S2**). *Exudation* is an input of organic carbon released from roots into the soil organic matter pool. Every live pool has a respiration rate (r) and a turnover or death rate (d). On consuming a C pool, a fraction of this pool always becomes faeces and enters the SOM pool except for the microbial pools, i.e. microbes and microbivores. SOM can be distributed in

different fractions, particulate organic matter (POM) and dissolved organic matter (DOM), which can gain relevance in the addition and simulation of other nutrient cycles and processes as leaching. However, here, as a first version of the model, we present a simplification using SOM as a uniform pool. The growth (G, g C m^{-3} day^{-1}) of a biomass pool (B, g C m^{-3}) is described according to Monod kinetic,

$$G = \sum_{n=1}^{N} (g_{max}\left(\frac{Sf_a}{K_s + S}\right)_n)B \qquad (9)$$

where g_{max} (g C g C^{-1} day^{-1}) is the maximal rate of growth, to which several modifiers are applied (see descriptions below). Substrate (S, g C m^{-3}) is the consumable pool, litter, SOM or biomass of soil organism (n), that consumer pool (B) can use but corrected by its available fraction (f_a). All fluxes (N) of consumed C from each S are summed. K_s (g C m^{-3}) is related to substrate quality, it gives the content required to get half the maximal growth. This is not related to the amount that will be consumed, because consumed C equals growth + faeces, but shows how dense the material needs to be 'found' by the consumer. Available fraction of a S to a consumer (as f_a) is calculated using the fraction from total porosity volume that is accessible for the consumer, by size, minus its fraction that is completely flooded or dry (see equations 14-16). This availability introduces the concept of physical recalcitrance, highlighting the role that soil structure plays affecting C fluxes in the soil, because SOM decomposition rates modelling use to rely on its chemical recalcitrance, from now on referred just as 'recalcitrance'. But physical recalcitrance has proven to be also relevant for the calculation of SOM decomposition rates (von Lützow *et al*., 2008), and soil matrix also affect other biotic interactions through the food web by this availability concept.

Rate of increase of a population of meso- or macrofauna depends on generation time (r, K strategies), age distribution of the population, different life stages. Models exist for only some soil fauna species (Osler and Sommerkorn, 2007; Chertov *et al*. 2017a). To offer a solution that can work for both the microbial biomass and the meso- and macrofauna, we use g_{max} as equal to the maximal rate of increase in biomass of any population, $dB/dt = g_{max}$ when resources are non-limiting and assuming the population structure is stable and optimal, equal to what is often stated as the intrinsic growth rate of a species (Birch, 1948).

The net rate of change of a biomass pool is the sum of growth (G), respiration (R) and turnover (death, D_t), and possibly predation (P_d), all in g C m^{-3} day^{-1}:

$$\frac{dB}{dt} = G - R - D_t - P_d \qquad (10)$$

R is a function of temperature, through respiration rate (r, g C g C^{-1} day^{-1}), and biomass, assuming the same temperature sensitivity as growth; this is somewhat different to how it is seen in many models where a food source is turned over with a specific efficiency. From a more faunal point of view, this makes sense: a food source is 'consumed'; the consumed material is partly excreted and partly assimilated and spent on respiration and growth (i.e. biomass formation).

$$R = rB \qquad (11)$$

While the death rate d (g C g C^{-1} day^{-1}) is constant.

$$D_t = dB \quad (12)$$

Predation depends on biomass of predator or microbivore and is calculated from the growth of the predator (G_{pred}, from equation 9) plus the fraction of the prey allocated to faeces (f_{faec}).

$$P_{d_{prey}} = G_{pred}\,(1 + f_{faec}) \quad (13)$$

Effect of H_2O

Drought or saturation of a pore leads to reduced availability of the C in the pore for its food web consumers. First, the overall effect of hydration is calculated as a modifier (m_{H2Otot}) in function of volumetric soil moisture (V) and pore volume (P_{vol}) (after Freytag and Luttich, 1985).

$$m_{H2Otot} = \begin{cases} 4\frac{V}{P_{vol}}\left(1 - \frac{V}{P_{vol}}\right) & \text{for } \frac{V}{P_{vol}} < 0.5 \\ 1 & \text{for } \frac{V}{P_{vol}} > 0.5 \end{cases} \quad (14)$$

The activity is always in the pores that are not water-logged, therefore the pore size class that is partially filled with water, and the pore size above that is assumed not yet completely dry (after Kuka, Franko and Rühlmann, 2007).

$$m_{H2O} = \frac{P_{volA}}{P_{volA} + P_{volW}} m_{H2Otot} \quad \text{for the pores partially filled,} \quad (15)$$

$$m_{H2O} = \frac{P_{volW}}{P_{volA} + P_{volW}} m_{H2Otot} \quad \text{for the pores one class above,} \quad (16)$$

where P_{volW} is the water filled pore volume and P_{volA} is the aerated pore volume. The availability (f_a) of a substrate to a consumer is defined by the inherent availability of the pore size to the consumer, multiplied with m_{H2O}. For surface litter these calculations are not possible since the surface litter is not in the soil matrix. However, on days without precipitation, litter humidity is assumed to be related to the soil humidity below, therefore the m_{H2O} calculated for the microbial biomass is used.

Simulating the variability in g_{max}

The maximum growth of biota is influenced by different environmental factors. Each one can lead to a modifier ($m \in [0, 1]$) on g_{max}. It is easy to change, add or turn off specific modifiers according to the soil studied. Here we present a modelling framework focused on abiotic controls of growth rates, but there is room for new add-ons as for example death rate modifiers as a density-dependent microbial turnover. While interaction processes affected by the demographic density of microbial communities (e.g. competition, space constraints) can play also a significant role controlling growth and decomposition rates and improve its modelling (Georgiou *et al*., 2017), our aim in this work is to link the key roles of fauna and soil structure in C cycle modelling, and together with the hydrology can simulate constraints in biotic interactions, which are also relevant controls in microbial growth and activity.

Simulating the effect of temperature (T)

To simulate the effect of T on growth rate through a temperature modifier (m_T), we use a Q10-shaped curve between maximum tolerable temperature (T_{max}) and minimum temperature for consumers activity (T_{min}), set as a default at 0°C (Franko, 1989), but unlike many models, we assume a plateau above the optimal temperature (T_{opt}).

$$m_T = \begin{cases} 0, & T < T_{min} \text{ or } T \geq T_{max} \\ Q^{(T - T_{opt})/10}, & T_{min} \leq T < T_{opt} \\ 1, & T_{opt} \leq T < T_{max} \end{cases} \quad (17)$$

However, temperature also increases respiration (R). To simulate this temperature effect, we assume the same Q10 function but without the plateau; in this way, when T is above the optimum, R increases while growth does not. At some point these lines will cross and cause a net reduction in biomass.

Effect of pH on growth

A good example of an optional effect is the effect of pH: for a system close to a threshold, simulating pH can be very important, assuming a good knowledge of the system. But for well-buffered systems, it is an unnecessary increase in complexity. g_{max} decreases at low pH for bacteria but increases for fungi (Rousk, Brookes and Bååth, 2009; Rousk et al., 2010; Rousk, Brookes and Bååth, 2010). For this example, we put the thresholds at 8 for fungi and 3 for bacteria, with precision of one decimal, inducing a 10 fold reduction in g_{max} for a change in pH of 1.

$$m_{pH} = 1/((pH - 8)10) \quad \text{for fungi if } pH \geq 8.1 \quad (18)$$

$$m_{pH} = 1/((3 - pH)10) \quad \text{for bacteria if } pH \leq 2.9 \quad (19)$$

In any other case for bacteria or fungi, $m_{pH} = 1$, and if m_{pH} goes above 1, it is replaced by 1. For engineer saprotrophs, their optimal g_{max} changes (becoming g_{maxEng}) with pH (Lavelle, Chauvel and Fragoso, 1995; Chertov *et al*., 2017b) according to the following equation:

$$g_{maxEng} = \begin{cases} 0, & \text{if } pH < 3 \\ \left(\frac{g_{max}}{2}\right)(pH - 3), & \text{if } 3 \leq pH < 5 \\ g_{max}, & \text{if } pH \geq 5 \end{cases} \quad (20)$$

Effect of recalcitrance and C:N on g_{max}

Overall consumption of an organism that can consume different pools is computed by simply adding them up. However, litter is not necessarily as ‘palatable’, depending on its C:N ratio if

not enough N, and on recalcitrance if low in energy, then it is needed to consume more litter, which is calculated through modifiers $m_{C:N}$ and m_{rec}, respectively. This is simulated by changing g_{max}. The equation for m_{rec} is not necessary and only important if enough data on litter quality is available or the users are interested into looking into the effects of changes in litter quality. The litter pool can be consumed by both bacteria and fungi, and of course also detritivores. Depending on the C:N ratio, the competition between these two is different; this is simulated by the g_{max} of the bacteria being more variable with C:N ratio. The sensitivity is described by the parameter $p_{mC:N}$, between 0 and 1.

$$\text{fungi: } m_{C:N_{fung}} = \min\left(1, \left(\frac{C:N_{fung}}{C:N_{lit}}\right)^{p_{mC:Nfung}}\right) \quad (21)$$

$$\text{bacteria: } m_{C:N_{bact}} = \min\left(1, \left(\frac{C:N_{bact}}{C:N_{lit}}\right)^{p_{mC:Nbact}}\right) \quad (22)$$

For litter recalcitrance (Rec_{lit}), a linear equation instead of a power is chosen so that decay of the recalcitrant litter is 0 if $p_{mRec} = 1$ and is unaffected if $p_{mRec} = 0$. The reason for choosing a different equation than above is that constrain of labile litter decomposition by C:N ratio should not completely stop decomposition but adjust the decomposition rate, while recalcitrant fraction of litter could remain almost constant for long periods. The following equations determine if the recalcitrant fraction of litter remains stable or if it is affected by decomposers partially or even totally.

$$\text{fungi: } m_{rec_{fung}} = \min(1, 1 - p_{mRec_{fung}} Rec_{lit}) \quad (23)$$

$$\text{bacteria: } m_{rec_{bact}} = \min(1, 1 - p_{mRec_{bact}} Rec_{lit}) \quad (24)$$

Adding up all these modifying effects on g_{max}

We assume a complete additivity of the effects, so the different modifiers on g_{max} are multiplied to get the overall effect, m_{tot} in equation 25. Another optional approach could be to use only the most limiting effect, setting m_{tot} equal to the lowest modifier and ignoring the rest.

$$m_{tot} = m_T \, m_{C:N} \, m_{rec} \, m_{pH} \, m_{H2O} \quad (25)$$

Not assimilated C

The reduction in a substrate equals the growth of the consumer plus the C that goes to faeces (excrements) and to respiration. Fraction to excrement (f_{faec}) is a parameter of the consumer and assumed constant. However, one consumes more and a larger fraction becomes faeces at a lower substrate quality, for the meso- and macro fauna, because microbes do not produce excrements; the sensitivity of f_{faec} to C:N ratio is expressed by the modifier m_{faec}. This is however only relevant for the detritivores and engineers (equation 26) that eat SOM and litter which can contain extremely variable amounts of nutrients; for the predators and herbivores we assume the variability is minimal. For the microbes, it was calculated as an effect on g_{max}.

$$f_{faecEff} = f_{faec} + m_{faec}\frac{C{:}N_{SOM} - C{:}N_{eng}}{C{:}N_{SOM}} f_{faec} \qquad (26)$$

Calculations regarding engineers

Soil changes made by engineer species depend on their body width, but in the model this is simplified using initial parameters for engineers' effects that must be chosen based on an average width (see **Table S2.5** in **Supplemental File S2**); the model then simulates their daily effects using their biomasses. Bioturbation is a function of engineer biomass (B_{eng}, g C m^{-3}), which calculates organic matter moving to deeper layers: litter moving (g C_{lit} g C_{eng}^{-1} day^{-1}) from litter layer to end of burrow, and SOM moving by mixing of soil between layers (g C_{SOM} g C_{eng}^{-1} day^{-1}). In this first version of the model, with only one soil layer, bioturbation works as a C output flow, but in future versions with more layers it could be upgraded to C flows between them.

Burrow volume (PV_B, l m^{-3}) is a function of engineer biomass (g C_{eng} m^{-3}) and the ratio of pore volume to engineer biomass (VE_{ratio}, l g C_{eng}^{-1}) but towards a maximum (PV_{Bmax}, l m^{-3}). On the other hand, burrow turnover (tPV_B) happens at a constant rate, with average burrow lifespan of 10 years; porosity decreases and burrows become mesopores.

$$PV_B = \max\left(\min\left(PV_{Bmax},\ VE_{ratio}B_{eng}\right), \left(PV_{macro} - PV_{textmacro}\right)\left(1 - tPV_B\right)\right) \qquad (27)$$

where PV_{macro} is the current pore volume of the macropores and $PV_{textmacro}$ is the textural porosity of the macropores (see in next section). This could be improved in future versions including perturbations as the possible effect of heavy rain.

Porosity calculations

The pore volume is distributed in five classes by pore size. Initial pore size distribution is given or measured as the total pore volume (PV, l m^{-3}) in each class. The link between aggregation and porosity is hard to quantify. Regelink *et al*. (2015a) showed for different soils that overall soil porosity is the sum of the textural porosity determined by the proportion of clay, sand and silt fractions and aggregation porosity. They conclude that micropores, which they define <9 µm, are mainly situated within the aggregates, while mesopores are situated between dry-sieved aggregates. While Regelink *et al*. (2015a) have shown that total micro and mesoporosity (<1000 µm) increases with total aggregate content, Grosbellet *et al*. (2011) have provided evidence that pores in the range 30 – 300 µm decrease with aggregation. Despite of the generally lower ranges for mesopores (9 – 1000 µm) described for soil physics (Lal and Shukla, 2004; Regelink *et al*., 2015a), here mesopores are assumed to be physically accessible to mesofauna body size (ca. 100 – 2000 µm), so we consider that mesopores ranging 30 – 1500 µm are a reasonable compromise. Based on that, we decided to hypothesize that aggregation increases bacterial and micro- porosity while decreasing mesoporosity. However, we want to emphasize that further experimental studies are needed to establish robust relationships between aggregation and pore size distribution.

Aggregates are not calculated as a pool, but the effect of aggregation is included in the calculation of porosity as described below. The following three porosities contribute to total porosity:

- Textural porosity (PV_{text}): measured or calculated from % clay and sand.
- Additional aggregation porosity (PV_{Ag}): all porosity in surplus of textural, can be estimated, for example from PTF (PedoTransfer Function) or calculated empirically from SOM and fungal biomass, i.e. mycorrhiza and other fungi, max 2% porosity extra (equation 29). Aggregation (Ag) is the fraction (0–1) of the SOM aggregated calculated as (based on the data from Malamoud *et al*., 2009):

$$Ag = \min\left(1, \frac{c\left(B_{fung} + B_{myc}\right)}{B_{SOM}}\right) \quad (28)$$

  with an empirical parameter c = 10. The aggregation porosity is then calculated as:

$$PV_{Ag} = k\, Ag\, B_{SOM} \quad (29)$$

  with k = coefficient (2 l g C^{-1} m^{-3}) based on empirical data (Regelink *et al*., 2015a; Regelink *et al*., 2015b).

- Bioporosity (PV_B): biopores created by engineers. Pore formation by engineers increases macroporosity, increasing soil layer density, but at the same time reduces mesoporosity as engineers push soil aside and produce casts that are denser than average soil. The relative importance of these two effects depends on the engineers' activity patterns, and is reflected by the parameter $f_{PV} \in (0, 1)$, which gives the fraction of the change in biopore volume that increases macroporosity. Therefore, the counterpart of the biopore volume (1 - f_{PV}) PV_B is 'compensated' by a decrease in mesoporosity.

Conceptually, the total soil porosity is then the sum of:

$$PV_{tot} = PV_{text} + PV_{Ag} + f_{PV}PV_B \quad (30)$$

In the model, pore volume is calculated for each pore size separately.

The volume of micropores (PV_{micro}) and bacterial pores (PV_{bact}) increases with increasing aggregation. Apart from creating additional porosity depending on the total amount of aggregated SOM (eq. 29), aggregation also increases the relative micro- and bacterial pore volume at the expense of (textural) mesoporosity (PV_{meso}), therefore not increasing total porosity. This effect is controlled by available pore space between mineral particles (i.e. textural mesoporosity) and we assume that half of this mesoporosity can be affected by aggregation. In both cases, we assume that the increase in porosity due to aggregation is divided equally among micropores and bacterial pores. The pore volume in different size classes is calculated as:

$$PV_{macro} = PV_{textmacro} + PV_B \quad (31)$$

$$PV_{meso} = PV_{textmeso} - \left(1 - f_{PV}\right)PV_B - \frac{Ag}{2}PV_{textmeso} \quad (32)$$

$$PV_{micro} = PV_{textmicro} + k\frac{Ag}{2}B_{SOM} + \frac{Ag}{4}PV_{textmeso} \quad (33)$$

$$PV_{bact} = PV_{textbact} + k\frac{Ag}{2}B_{SOM} + \frac{Ag}{4}PV_{textmeso} \quad (34)$$

Volume of inaccessible pores is assumed to be constant and equal to $PV_{textinac}$.

These changes are calculated daily to give a dynamic feedback to the hydrology and to the distribution of each C source among pore classes, affecting its availability.

Leaching

Water leaving one soil layer (n) is moved to the layer below (n+1). Dissolved organic and inorganic compounds are a complex matter to simulate since they are strongly dependent on the pH and the mother-material, i.e. clay and Ca rich or not. Nonetheless, in many systems simulating leaching of N and DOM is highly relevant. Unless better data are available, we suggest the following, semi-empirical method:

DOM can be simulated in relation to the CO_2 released as total respiration (R_{tot}, g C m^{-3} day^{-1}) based on the consideration that high 'activity' in the soil is related to high R_{tot}. This is calculated as a fraction (f_{DOM}) of R_{tot} entering the DOM pool, similar to the concepts used in the LIDEL model, in addition to the directly exuded DOM (C_{Exud}, g C m^{-3} day^{-1}) which is an input (from data or a vegetation model). Assuming a short half-life of DOM and semi-empirically, because daily concentration is not 'equal' to daily production (DOM_p, g C m^{-3} day^{-1}) but is linearly related to the daily production, we consider:

$$DOM_p = C_{Exud} + f_{DOM} R_{tot} \quad (35)$$

DOM has a short half-life but the dissolution is even faster (hours). We assume the daily concentration is in equilibrium between dissolved and adsorbed (DOM_{ad}) depending on adsorption coefficient K_D of the soil (m^3 kg^{-1} soil). Similar to the modelling in Orchidee-SOM (Cammino-Serrano *et al*., 2018) we assume:

$$DOM_{ad} = K_D \, DOM \quad (36)$$

with K_D depending on the minerals and pH. More clay (fClay fraction) means less mobile DOM, and lower pH is also a cause of less mobile DOM.

$$K_D = a_{KD} - b_{KD} pH + c_{KD} \, fClay \quad (37)$$

with values 0.001226, 0.000212 and 0.00374 respectively for a_{KD}, b_{KD} and c_{KD}, from Cammino-Serrano *et al*. (2018).

DOM leaching is calculated as the volume of water leaching to a lower layer multiplied with the concentration of dissolved DOM.

Calculation order

The sequence of function sets used by the model to calculate all carbon fluxes and ecosystem changes is as follows:

a) Calculate the pore size fractions in 5 classes and the associated pore surface areas
b) Calculate the water volume of the relevant pore size
c) Use the precipitation leaching to calculate DOM leaching

d) Calculate for each biota group the accessibility of each of the pools it consumes
e) Calculate the g_{max} depending on temperature, H_2O, C:N, pH and recalcitrance
f) Solve the 12 differential equations for increase/decrease of all C pools
g) Update all C pools
h) Calculate the new C:N and recalcitrance of each pool
i) Calculate engineering effect
    a. Update macropores
    b. Update SOM from bioturbation
j) Calculate other changes in pore size distribution from weather or management

The KEYLINK core model consists of steps d to i; steps a, b, c and j are add-ons that could be replaced by other models (e.g. water flow model) coupled to KEYLINK. Steps a-c are used to calculate the distribution of porosity between the pore classes, the hydrology and daily soil water content (distributed among pore classes), and then step d calculates how that is affecting the availability of each C source to its consumers. That couples soil structure and hydrology with trophic interactions, allowing the resolution of differential equations for C fluxes.

Model coding and output

KEYLINK consists of a relatively limited, freely downloadable Python code (available at: https://github.com/Plant-Root-Soil-Interactions-Modelling/KEYLINK). Each of the modifiers on growth, i.e. temperature, pH, H_2O, recalcitrance and C:N, as well as the primal shape of the growth equations can be adapted towards specific questions or ecosystems. The inputs in the current version are read from data-files but are easy to link to a mechanistic model. The output of the current version consists of all daily C pools as well as the main C fluxes. KEYLINK is also available as a stand-alone executable model, allowing it to be called from models in other languages. A single run of ten years could take less than one minute (depending on computing power). In this version, the average results over one hundred runs are calculated but also all daily outputs of each run are saved.

Model parameterization

The first version of KEYLINK model has been parameterized for a Scots pine forest stand situated in Brasschaat, in the Campine region in Belgium (51°18’ N and 4°31’ E) but without modelling the complete forest ecosystem (to simplify the interpretation of the results from KEYLINK). The goal of this parameterization was a model verification, not a model application for which a complete integration with an aboveground model or detailed above ground data would be necessary.

The soil of the Brasschaat forest is sandy but with high ground water table, so trees are generally not water-limited, but the topsoil is often dry. The soil is acidic (pH 3.5). The trees were planted around 1930 and formed a rather sparse vegetation in 1999, with leaf area index (LAI) ranging from 2.1 to 2.4.

For this model run, we used the following input data from the stand (**Table 1**). In this case, we did not use measured or modelled growing trees but constant input of aboveground and

belowground litter (measured value). The top 90 cm of soil from the Brasschaat forest was analyzed in 1999 by Janssens *et al*. Earthworm biomass, used in this case as an example of ecosystem engineers, is extremely low due to the low pH, it was not measured since 1993 by Muys, but these data are used since there is no reason to expect there was a marked change.

Data availability on soil pools, biology and functioning is generally low, and it is currently not possible to find a dataset describing in detail, and with small error margins, the temporal evolution of all different soil biological compartments and SOM pools. Available data are often incomplete, or based on rough estimates, e.g. from semiquantitative DNA analysis for microbial abundance in soils. To deal with this issue, a quite pragmatic approach combining different estimates from different sources is appropriate for most datasets where the soil is not the key focus, but a means to improve the simulation of an ecosystem.

The daily loss of water by evapotranspiration was calculated using an equation for potential evapotranspiration based on Thornthwait (1948) in this study.

Model calibration

Once the model is parameterized for an ecosystem, the next step is to optimize that model, calibrating the fit of its simulations to the ecosystem data. The optimization included in the KEYLINK model follows a Bayesian procedure (Van Oijen, Rougier and Smith, 2005; Van Oijen, 2008).

A pragmatic assumption is that the starting values of the C pools (including the soil fauna initial biomasses) are at steady state for a given date (most often spring or summer, it would be unrealistic to keep the values constant through the year as they fluctuate with climate). The simplest calibration of any ecosystem can be done by assuming these 11 carbon pools (litter, SOM and the 9 functional groups in food web) need to be stable over the simulated years, e.g. for 9 years that gives us 99 data points by taking the same value for each C pool every year (**Table 2**). Initial litter, SOM and biomasses of bacteria, fungi and engineers were taken from the references cited in **Table 1**. For other C pools, data were estimated using measured data for previous C pools and similar proportions between C pools as in the Swedish pine forest in Persson *et al*. (1980); predator biomass was assumed to be the 20% of all biomass in their consumed C pools. Errors were assumed as a percentage of biomass, 10% for predators, 12.5% for litter and SOM, and 20% for the rest C pools.

It is common to apply a correction ("burn-in") deleting part of the posterior, e.g. the first half of the runs, to avoid the effect of the starting distribution (Gelman and Shirley, 2011). For this calibration, a sample of the last one hundred accepted parameter vectors was taken from the posterior distribution, and it was used for all further model runs, so every run was performed with 100 different parameter sets.

Input parameters of species

The KEYLINK model framework is conceptualized as an adaptable framework. Each user needs to determine for their specific site and questions the main drivers and pools required. Depending on the dataset, it is in general better to use less pools and equations if sparse data are available. Moreover, KEYLINK is not a soil fauna model and was not designed to simulate specific soil

fauna species in detail. The soil fauna groups used consist of a wide range of species, for which average data are used. For a description of the species categories, we refer to the review on the KEYLINK concepts (Deckmyn *et al*., 2020).

Microbes and meso-macro fauna have a temperature curve using an optimum, minimum and maximum temperature. Each soil biota group also has its own maximum growth rate, C:N ratio, respiration rate and size. Death rate (d) is the inverse of turnover, mostly given in days. In **Supplemental File S1** we briefly review main input parameters. We propose setting K_s, the concentration of the food source at which growth rate is half the maximum, equal to the existing concentrations for all meso- and macro fauna, so assuming growth could double at unlimited food source. But for microbial biomass the difference between growth of bacteria on a petri-dish unlimited in nutrients compared to field data of soil microbes clearly indicates that g_{max} in the soil is not comparable to laboratory data; if such data of g_{max} are used, the K_s should be increased considerably.

Calibration for Brasschaat pine forest

We show here the results from a calibration towards data measured and assumed, using proportions between fauna groups in Persson *et al*. (1980), in the Brasschaat Scots pine stand in Belgium. This forest stand is relatively well described in many publications concerning the trees and the total ecosystem fluxes, but less concerning the soil and very little was measured on soil fauna. Therefore, the partially assumed data refers to a hypothetical ecosystem that does not fully fit with reality in Brasschaat forest. We use this forest as an example of how the KEYLINK model can be used to improve our understanding of the system even when detailed soil faunal data are limited.

The parameters $g_{max,}$ K_s and r are linked (increasing g_{max} has a similar effect to decreasing K_s or r). However, g_{max} or r ranges can be found in literature relatively easily. Therefore, we use fixed values for K_s (see **Supplemental File S2**) and parameterize g_{max} within the known limits. In this way, the number of parameters to be calibrated is 9, which is a reasonable number for most cases where limited data to calibrate towards are available. Of course, a user could decide to optimize more parameters if more data are available. A useful ‘rule of the thumb’ is limiting the number of parameters to the square root of the number of calibration data available (Jörgensen, 2009), which means we can get a reasonable result for nine parameters assuming 81 data points.

In our case, no measurements of growth rates were available and information in the literature was scant. Therefore, we deliberately defined wide ranges for the prior values of each parameter to cover all the possible values found in the literature (Chuine, 2000; Linkosalo, Lappalainen and Hari, 2008). For species for which no prior parameter information was available, we assumed parameter values equal to the mean value of the range. The initial uncertainty of each parameter is quantified in terms of a prior probability distribution with lower and upper bounds. Because of lack of detailed knowledge, we assumed the distribution as uniform and non-correlated.

The g_{max} values were optimized using the prior range for g_{max} (**Table 3**). The data used to calibrate against were chosen to give a ‘standard’ procedure, so limited to biomass of the different C pools. Including all available data s.a. soil respiration, soil humidity could improve the run for Brasschaat, but would not be a representative run for the model. Other parameter

settings, e.g. sensitivity to C:N and recalcitrance, were based on model runs of the Brasschaat site by Deckmyn *et al*. (2011).

We ran the model for the time period 1999-2008, because this was the period in which the forest was still clearly dominated by Scots pine; since then a transition to more deciduous trees has been taking place. We calibrated towards stable C pools over the ten years for all C pools, with an allowed error margin of 20% for all faunal pools, except 10% for predators, and 12.5% for litter and SOM. Daily climate data (temperature and precipitation) were used. The full range of input data can be found in **Supplemental File S2**, except climate data, which can be downloaded with the model. Choosing to calibrate towards one or more pools can yield different results, and it depends on the end-user's goal which calibration is preferred.

Model evaluation

Although coupling KEYLINK to real or simulated data of the aboveground ecosystem would yield more realistic results, in this exercise we used KEYLINK as a stand-alone model with quite constant input (e.g. litter, plant water uptake) to minimize the feedback effects and give a clear view on the model behaviour. This is a model evaluation, not a full model validation.

After calibration to the Brasschaat dataset, a set of scenarios was performed to evaluate the model: I. Basic results; II. Sensitivity to initial soil structure; III. Changing initial litter C:N ratio; IV. Changing initial litter recalcitrance; V. Changing soil pH; VI. Excluding predators.

Scenario I was done with the reference input parameters (**Supplemental File S2**), and used as a basal one to be compared with the other five alternative scenarios: scenario II with higher clay content in the soil (clay 15%); scenario III with lower litter C:N ratio (40); scenario IV with lower litter recalcitrance (20%); scenario V with higher pH (5.9); and scenario VI without predators by setting its initial biomass to 0 ($B_{pred} = 0$).

In each one of the five alternative scenarios, input parameters were the same than in the basal scenario, except for the parameter changed to generate the new scenario (see **Supplemental File S2**). All the six scenarios were run 100 times using a sample of 100 parameter vectors from the posterior distribution of the calibration, consisting each run in a simulation of 10 years at a daily time-step (3653 days). Then, averages of biomasses were calculated for each C pool among the 100 simulations of 10 years, for each scenario, comparing the effects of disturbances on average values.

Results

Calibration

The model was run ca. 100000 times with different parameter settings sampled from the prior parameter distribution. A sample of the posterior distribution was taken with the last 100 accepted parameter vectors for g_{max} (**Table 4**).

The optimization showed the link between the different groups of soil biota, e.g. a high g_{max} for bacteria was coupled to a high g_{max} for bacterivores. The alternative five scenarios compared to

the basal one can show very different results concerning specific C pools (**Table 5**). Running the model 100 times using the sample of the g_{max} values resulted in predictions with a quite wide range (**Table 6**).

Basic results

Mycorrhiza, herbivores and detritivores are relatively uncoupled, though influenced by predators, and follow the yearly climate curves. The bacterial and fungal biomasses are very strongly linked. The high g_{max} of bacteria allows steep peaks, which are generally followed by peaks in bacterivore biomass. As we used constant litter input into the soil and used a calculated constant fraction of potential evapotranspiration as water uptake from the soil, it cannot be expected that these results follow the normal annual trends in fluctuations of those fluxes. This can, at least partially, explain the relatively low bacterial biomass found in our results, since the bacteria would profit most from a rapidly changing environment, but under some unrealistic simulated conditions fungi could be displacing bacteria by competitive exclusion. For more realistic simulations the model can be coupled with other models that give that information as outputs, or with measured datasets.

All C pools tend to reach some stability after the first years, suggesting the model is well-balanced; however, stability values seem to be more sensitive to changes in g_{max} parameters for some pools (e.g. SOM). The set-up of the model, where we only calibrate the faunal g_{max}, does not allow calibration towards different ratio of litter and SOM decay. This depends on the uncalibrated parameter fragmentation, the sensitivity to recalcitrance, but also the temperature used for the litter and SOM. Here we used the same temperature, while in reality it would be expected that there would be certain differences in mean temperatures or their variability at different depths.

There was a high variability within the 100 simulations of each scenario, e.g. basal scenario (**Fig. 3**), calling into question the reliability of predictions. We suggest here some theoretical predictions based on this example, but it is clear that for a realistic application it will be necessary to improve the calibration, by using more detailed data or by linking the model to a vegetation model. This highlights the relevance of developing databases including enough details for the key parameters of the different parts of the soil system (i.e. soil structure, hydrology, food webs).

An overview of all C pools under the different simulation scenarios (**Fig. 4**) shows how changing one input parameter at a time influences the results. It must be taken into account that KEYLINK was run as a stand-alone model, which can explain why some of the resulting outputs seem not very realistic; linking it to a model or more detailed data of the aboveground ecosystem would greatly influence the results, but would not allow clear interpretation of the model functioning due to feedbacks. Since our goal here was to introduce the belowground model itself rather than a realistic application to a particular case, we chose to avoid those feedbacks with other parts of the ecosystem that are required for more realistic simulations. To further elucidate these effects and to show some of the potential outputs the model can give, we show a few of the most interesting fluxes (**Table 7**).

The simulated scenarios showed that increasing clay content (i.e. changing initial soil structure) resulted in an increase in water content (**Fig. 5**) and a decrease in litter and SOM decay while fungal/bacterial ratio decreased. In fact, this scenario caused the largest change in soil water

content, showing the sensitivity of the system to initial soil structure and its crucial role for soil hydrology. On the other hand, lowering litter recalcitrance or C:N ratio resulted in an increase in microbial biomass, mainly fungi, which caused an increase in litter decay, while SOM decay did not show a clear change. The scenario with higher pH allowed engineer species (and predators with them) to increase in biomass, altering the soil structure with an increase in macroporosity, which caused a clear decrease in SOM stabilization. Finally, the exclusion of predators totally changed the soil porosity and the trophic interactions along the food web, causing the largest increase in the decay of SOM and litter, which highlights the crucial role of predators in the regulation of the soil C cycle.

Discussion

The Brasschaat forest is sandy, with low pH and recalcitrant litter; as expected, this is an environment not suited to earthworms. The model correctly simulated extremely low values of engineer biomass. Increasing the pH increased the engineers pool, e.g. earthworms population, but this remained too low to have a significant impact on the system. This is quite realistic as neither litter quality nor soil quality are ideal for earthworms. Obviously, to calibrate the specific parameters concerning earthworms the Brasschaat forest is not an ideal site.

The high variability observed in some populations could make them appear more unstable than what can be expected in reality, and it is indeed expectable that a more realistic application of the model will yield different results. However, there are also empirical evidences that some populations, e.g. microbial biomasses, can be very unstable depending on hydrology (Blackwell *et al*., 2010; Zhao *et al*., 2010), showing short-term spikes in biomass as those observed in the presented simulations, in response to predicted changes in water availability.

Changes in litter quality (i.e. in C:N ratio and recalcitrance) caused small differences between scenarios, being the simulations from those two alternative scenarios quite similar to the basal scenario. Despite the apparent increase in litter decay for lower litter recalcitrance and for lower C:N ratio, as mentioned before, the high variability among simulations (**Fig. 3**) suggests that differences between scenarios caused by changes in litter quality are negligible. It is well known that litter quality is one of the main factors controlling litter decomposition rates (Zhang *et al*., 2008; García-Palacios *et al*., 2013), and the observed trend fits with experimental evidence, but clearer effects of changes in litter quality are expected. Therefore, these simulations should be tested again, particularly with the model linked to a vegetation model, and, if necessary, improvements should be added.

Changes in textural porosity, on the other hand, showed clearer effects on soil processes. The scenario changing clay content to increase textural porosity caused the highest impact on hydrology, increasing soil water content (E in **Figure 5**), which led to a decrease in SOM availability, decreasing microbial community biomasses and, therefore, decreasing the decomposition of SOM and litter. However, changes in C pools were not as clear as in hydrology, mainly due to the high uncertainty in calibration results.

The scenario without predators showed the most interesting results because the interactions between the different food web parts are apparent, and it showed a high contrast with the basal

scenario. Exclusion of predators, setting the starting biomass of that pool at 0, showed how the model tracks its crucial role in the ecosystem (**Fig. 4**). Predators produce a top-down trophic cascade on the food web, e.g. on herbivores and roots. Microbial decay is reduced as fungi and bacteria are consumed by the increased populations of bacterivores and fungivores. Despite of the decrease in bacteria and fungi, SOM and litter were also lower without predators. This could be explained by an increase in engineer populations with implications also on the soil matrix. Overall the model successfully tracked soil food web dynamics and also their interactions with soil porosity. The effect of larger soil predators (e.g. Araneae, Carabidae, Formicidae) slowing down SOM decomposition and enhancing its stabilization has been previously found in experiments (Kajak, 1995), as well as mycorrhiza effect on porosity by making aggregates (Siddiky *et al*., 2012). Those trophic cascade effects over SOM stabilization depends on environmental conditions such as rainfall, with predation on microbivores reducing litter decomposition rates in more humid sites, while reductions in rainfall could lead to a shift in that trend with predation on microbivores indirectly increasing litter decomposition rates (Lensing and Wise, 2006). The predicted effects of predator exclusion increasing SOM and litter decay have been found in different ecosystems, e.g. grasslands (Kajak, 1997) and forests (Lawrence and Wise, 2000), but contrasting results (e.g. Kajak, 1995; Lawrence and Wise, 2004) suggest that it is not a general pattern. Considering the high rainfall conditions at the modelled Belgian forest, and according to the suggested trends based on experimental research, soil predators feeding on microbivores at Brasschaat forest could promote SOM stabilization, which would fit with the simulated scenarios. Therefore, KEYLINK model seems to fit with the expected food web and C dynamics, and could serve to improve the biogeochemical cycles modelling, as is needed for larger scale predictions (Grandy *et al*., 2016), by coupling it with other ecosystem models.

Hydrology is influenced by aggregation and by macropore formation by ecosystem engineers. The increased macroporosity increases infiltration rate with reduced water-logging and runoff. Predators have a clear indirect effect on soil porosity by consuming engineer species, and also microbivore species, which leads to changes in soil hydrology (**Fig. 5**). Variations in bacterial pore and micropore volume are positively correlated, while mesopore variations are negatively correlated with both; the higher volume in mesopores, the lower in the two other classes, and the faster the water drains from the soil layer. That is what we can expect to happen in real soils, so the model seems to simulate appropriately those dynamics. The increase in macro and mesoporosity volumes without predators, so with higher engineers, resulted in a decrease of soil water content of 9.26 % (increasing the pore aeration), and under those conditions the availability of SOM and litter for bacteria and fungi could be increased, explaining why SOM and litter are lower even with lower bacteria and fungi. This highlights the role of hydrology on trophic cascade processes, which can be enhanced or reduced by water distribution through the soil matrix (Erktan, Or and Scheu, 2020), and also the relevance of considering how climate change effects on soil structure, hydrology and food web interactions (particularly trophic cascades) can affect microbial communities (Thakur and Geisen, 2019) and, therefore, litter and SOM decomposition.

The aim of this study was to present a first version of a new concept model that hopefully will serve to challenge current state-of-the-art soil modelling. But we are aware that to do that we will need to improve the calibration of the model in the future, using more complete databases that take into account all the elements needed to calibrate KEYLINK, which, on the other hand, are

currently extremely scarce. By presenting this concept model that challenges the current way of simulating soil biochemical cycling, we hope to stimulate that future studies will also be designed to take into account the pools and functional groups needed to calibrate KEYLINK.

Conclusions

KEYLINK is a relatively simple, fast and easy to modify soil model that can be used as a stand-alone model to understand soil systems, or linked to detailed aboveground data/models to predict SOM turnover. Model evaluation showed that KEYLINK is capable of simulating properly not only the soil food web and C pools dynamics, but also how they interact with soil porosity and hydrology, which is one of the main goals of this new model. The results from the evaluation scenarios showed that SOM turnover is driven not only by microbial biomass, but also by soil structure and hydrology. Moreover, microbial biomass is strongly regulated by the presence/absence of the other soil fauna. Especially the effects of the predators and the ecosystem engineers are extremely significant for our understanding of soil functioning. Furthermore, since management can differentially affect the larger soil fauna, KEYLINK can be of great use to investigate potential effects of management changes on soil SOM, nutrient turnover and hydrology.

This model shows degradability of SOM can be adequately simulated from accessibility in relation to pore space instead of the existing concepts of slow and fast pools. This allows a closer link to the soil structure and soil fauna which we consider closer to the actual, and follows the concepts as first described by Kuka, Franko and Rühlmann (2007), but applied here in a wider framework and including the hydrology.

For a full validation or better calibration of the model, datasets are required including basic data on the aboveground, e.g. litter input, water uptake, root growth and turnover, in combination with relatively detailed data on soil structure, i.e. pore size distribution, and hydrology and soil biota, e.g. biomass of bacteria, fungi, mycorrhizal fungi and main meso-and macrofauna. All these data are available, but very seldom at one site as most studies are focused on one or other aspect of soil science.

In conclusion, KEYLINK is a step towards a new generation of ecosystems models that include functional diversity, trophic structures and ecological processes as important factors shaping soil/ecosystem carbon and water cycling. Future versions, fed by more detailed data, will need to be developed in order to improve our current predictive capacity.

Acknowledgements

The authors express their gratitude to all the people who have contributed to the BioLink and KEYSOM COST Actions, whose work contributed also to the development of the KEYLINK model.

References

777 Birch, L. (1948). The intrinsic rate of natural increase of an insect population. *The Journal of*
778 *Animal Ecology, 17*(1), 15-26.

779 Blackwell, M. S. A., Brookes, P. C., de La Fuente-Martinez, N., Gordon, H., Murray, P. J.,
780 Snars, K. E., Williams, J. K., Bol, R., and Haygarth, P. M. (2010). Phosphorus
781 solubilization and potential transfer to surface waters from the soil microbial biomass
782 following drying–rewetting and freezing–thawing. In *Advances in agronomy* (Vol. 106,
783 pp. 1-35). Academic Press.

784 Camino-Serrano, M., Guenet, B., Luyssaert, S., Ciais, P., Bastrikov, B., De Vos, B., Gielen, B.,
785 Gleixner, G., Jornet-Puig, A., Kaiser, K., Kothawala, D., Lauerwald, R., Peñuelas, J.,
786 Schrumpf, M., Vicca, S., Vuichard, N., Walmsley, D., and Janssens, I. A. (2018).
787 ORCHIDEE-SOM: modeling soil organic carbon (SOC) and dissolved organic carbon
788 (DOC) dynamics along vertical soil profiles in Europe. *Geoscientific Model*
789 *Development, 11*(3), 937-957.

790 Campbell. E. E., Parton, W. J., Soong, J. L., Paustian, K., Hobb, N. T., and Cotrufo, M. F.
791 (2016). Using litter chemistry controls on microbial processes to partition litter carbon
792 fluxes with the Litter Decomposition and Leaching (LIDEL) model. *Soil Biology and*
793 *Biochemistry, 100*, 160-174.

794 Campbell, E. E., and Paustian, K. (2015). Current developments in soil organic matter modeling
795 and the expansion of model applications: a review. *Environmental Research Letters,*
796 *10*(12), 123004.

797 Chertov, O., Komarov, A., Shaw, C., Bykhovets, S., Frolov, P., Shanin, V., Grabarnik, P.,
798 Priputina, I., Zubkova, E., and Shashkov, M. (2017a). Romul_Hum - A model of soil
799 organic matter formation coupling with soil biota activity. II. Parameterisation of the soil
800 food web biota activity. *Ecological Modelling, 345*, 140-149.

801 Chertov, O., Shaw, C., Shashkov, M., Komarov, A., Bykhovets, S., Shanin, V., Grabarnik, P.,
802 Frolov, P., Kalinina, O., Priputina, I., and Zubkova, E. (2017b). Romul_Hum model of
803 soil organic matter formation coupled with soil biota activity. III. Parameterisation of
804 earthworm activity. *Ecological Modelling*, *345*, 140-149.

805 Christian, P. R., and Casella, G. (1999). Monte Carlo statistical methods. Springer. New York.

806 Chuine, I. (2000). A unified model for budburst of trees. *Journal of Theoretical Biology, 207*(3),
807 337-347.

808 Cotrufo, M. F., Wallenstein, M. D., Boot, C. M., Denef, K., and Paul, E. (2013). The Microbial
809 Efficiency-Matrix Stabilization (MEMS) framework integrates plant litter decomposition
810 with soil organic matter stabilization: do labile plant inputs form stable soil organic
811 matter? *Global Change Biology, 19*, 988-995.

812 Deckmyn, G., Campioli, M., Muys, B., and Kraigher, H. (2011). Simulating C cycles in forest
813 soils: Including the active role of micro-organisms in the ANAFORE forest
814 model. *Ecological Modelling, 222*, 1972-1985.

Deckmyn, G., Meyer, A., Smits, M. M., Ekblad, A., Grebenc, T., Komarov, A., and Kraigher, H., (2014). Simulating ectomycorrhizal fungi and their role in carbon and nitrogen cycling in forest ecosystems. *Canadian Journal of Forest Research, 44*(6), 535-553.

Erktan, A., Or, D., and Scheu, S. (2020). The physical structure of soil: Determinant and consequence of trophic interactions. *Soil Biology and Biochemistry*, 107876.

Filser, J., Faber, J. H., Tiunov, A. V., Brussaard, L., Frouz, J., De Deyn, G., Uvarov, A. V., Berg, M. P., Lavelle, P., Loreau, M., Wall, D. H., Querner, P., Eijsackers, H., and Jiménez, J. J. (2016). Soil fauna: key to new carbon models. *Communications in Soil Science and Plant Analysis, 2*, 565-582.

Franko, U. (1989). C-und N-Dynamik beim Umsatz organischer Substanzen im Boaen. Diss., Berlin, Akad. Landwirtsch.

Freytag, H. E., and Luttich, M. (1985). Zum Einfluß der Bodenfeuchte auf die Bodenatmung unter Einbeziehung der Trockenraumdichte. *Archiv für Acker-und Pflanzenbau und Bodenkunde*, *29*(8), 485-492.

García-Palacios, P., Maestre, F. T., Kattge, J., and Wall, D. H. (2013). Climate and litter quality differently modulate the effects of soil fauna on litter decomposition across biomes. *Ecology letters*, *16*(8), 1045-1053.

Gaublomme, E., De Vos, B., and Cools, N. (2006). An indicator for microbial biodiversity in forest soils. *Brussels (BE): Instituut voor Natuur-en Bosonderzoek (INBO). R.*

Geisen, S., Briones, M. J., Gan, H., Behan-Pelletier, V. M., Friman, V.P., de Groot, G. A., Hannula, S. E., Lindo, Z., Philippot, L., Tiunov, A. V., and Wall, D. H. (2019). A methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry, 136*, 107536.

Gelman, A., and Shirley, K. (2011). Inference from simulations and monitoring convergence. In: Brooks, S., Gelman, A., Jones, G., and Meng, X. L. (eds.), *Handbook of Markov Chain Monte Carlo, 6*, 163-174. CRC press, USA.

Georgiou, K., Abramoff, R. Z., Harte, J., Riley, W. J., and Torn, M. S. (2017). Microbial community-level regulation explains soil carbon responses to long-term litter manipulations. *Nature Communications, 8*(1), 1-10.

Grandy, A. S., Wieder, W. R., Wickings, K., and Kyker-Snowman, E. (2016). Beyond microbes: Are fauna the next frontier in soil biogeochemical models? *Soil Biology and Biochemistry, 102*, 40-44.

Grosbellet, C., Vidal-Beaudet, L., Caubel, V., and Charpentier, S. (2011). Improvement of soil structure formation by degradation of coarse organic matter. *Geoderma*, *162*(1-2), 27-38.

Horemans, J., Roland, M., Janssens, I., and Ceulemans, R. (2017). Explaining the inter-annual variability in the ecosystem fluxes of the Brasschaat Scots pine forest: 20 years of eddy flux and pollution monitoring. In: *EGU General Assembly Conference Abstracts, 19*, 10402.

Janssens, I. A., Sampson, D. A., Cermak, J., Meiresonne, L., Riguzzi, F., Overloop, S., and Ceulemans, R. (1999). Above- and belowground phytomass and carbon storage in a Belgian Scots pine stand. *Annals of Forest Science 56*, 81-90.

Janssens, I. A., Sampson, D. A., Curiel Yuste, J., Carrara, A., and Ceulemans, R. (2002). The carbon cost of fine root turnover in a Scots pine forest. *Forest Ecology and Management, 168*(1-3), 231-240.

Jones, C. G., Lawton, J. H., and Shachack, M. (1994). Organisms as ecosystem engineers. *Oikos 69*, 373-386.

Jörgensen, S. E. (2009). *Ecological modelling: an introduction*. WIT press. ISBN 978-1-84564-408-6

Kajak, A. (1995). The role of soil predators in decomposition processes. *European Journal of Entomology, 92*(3), 573-580.

Kajak, A. (1997). Effects of epigeic macroarthropods on grass litter decomposition in mown meadow. *Agriculture, ecosystems & environment*, *64*(1), 53-63.

Kuka, K., Franko, U., and Rühlmann, J. (2007). Modelling the impact of pore space distribution on carbon turnover. *Ecological Modelling, 208*(2-4), 205-306.

Lal, R., and Shukla, M. K. (2004). *Principles of soil physics*. CRC Press.

Lavelle, P., Chauvel, A., and Fragoso, C. (1995). Faunal activity in acid soils. In *Plant-Soil Interactions at Low pH: Principles and Management* (pp. 201-211). Springer, Dordrecht.

Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jiménez, J. J., McKey, D., Mathieu, J., Velasquez, E., and Zangerlé, A. (2016). Ecosystem engineers in a self-organized soil: a review of concepts and future research questions. *Soil Science, 181*(3/4), 91-109.

Lawrence, K. L., and Wise, D. H. (2000). Spider predation on forest-floor Collembola and evidence for indirect effects on decomposition. *Pedobiologia*, *44*(1), 33-39.

Lawrence, K. L., and Wise, D. H. (2004). Unexpected indirect effect of spiders on the rate of litter disappearance in a deciduous forest. *Pedobiologia*, *48*(2), 149-157.

Lensing, J. R., and Wise, D. H. (2006). Predicted climate change alters the indirect effect of predators on an ecosystem process. *Proceedings of the National Academy of Sciences*, *103*(42), 15502-15505.

Linkosalo, T., Lappalainen, H., and Hari, P. (2008). A comparison of phenological models of leaf bud burst and flowering of boreal trees using independent observations. *Tree Physiology, 28*(12), 1873-1882.

Malamoud, K., McBratney, A. B., Minasny, B., and Field, D. J. (2009). Modelling how carbon affects soil structure. *Geoderma, 149*, 19-26.

Mambelli, S., Bird, J. A., Gleixner, G., Dawson, T. E., and Torn, M. S. (2011). Relative contribution of foliar and fine root pine litter to the molecular composition of soil organic matter after in situ degradation. *Organic Geochemistry, 42*, 1099-1108.

Muys, B. (1993). A synecological evaluation of the earthworm activity and litter decomposition in Flemish forests in the context of sustainable forest management. D. Phil. Thesis, University of Ghent.

Osler, G. H., and Sommerkorn, M. (2007). Toward a complete soil C and N cycle: incorporating the soil fauna. *Ecology, 88*(7), 1611-1621.

Persson, T., Bååth, E., Clarholm, M., Lundkvist, H., Soderstroem, B. E., and Sohlenius, B. (1980). Trophic structure, biomass dynamics and carbon metabolism of soil organisms in a Scots pine forest. *Ecological Bulletins, 32*, 419-459.

Regelink, I. C., Stoof, C. R., Rousseva, S., Weng, L., Lair, G. J., Kram, P., Nikolaidis, N. P., Kercheva, M., Banwart, S., and Comans, R. N. J. (2015a). Linkages between aggregate formation, porosity and soil chemical properties. *Geoderma 247*, 24-37.

Regelink, I. C., Weng, L., Lair, G. J., and Comans, R. N. J. (2015b). Adsorption of phosphate and organic matter on metal (hydr)oxides in arable and forest soil: a mechanistic modelling study. *European Journal of Soil Science, 66*(5), 867-875.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R., and Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *The ISME journal*, *4*(10), 1340-1351.

Rousk, J., Brookes, P. C., and Bååth, E. (2009). Contrasting soil pH effects on fungal and bacterial growth suggest functional redundancy in carbon mineralization. *Applied and Environmental Microbiology*, *75*(6), 1589-1596.

Rousk, J., Brookes, P. C., and Bååth, E. (2010). Investigating the mechanisms for the opposing pH relationships of fungal and bacterial growth in soil. *Soil Biology and Biochemistry*, *42*(6), 926-934.

Saxton, K. E., Rawls, W., Romberger, J. S., and Papendick, R. I. (1986). Estimating generalized soil-water characteristics from texture 1. *Soil Science Society of America Journal, 50*(4), 1031-1036.

Schmidt, M. W. I., Torn, M. S., Abiven, S., Dittmar, T., Guggenberger, G., Janssens, I. A., Kleber, M., Kögel-Knabner, I., Lehmann, J., Manning, D. A. C., Nannipieri, P., Rasse, D. P., Weiner, S., and Trumbore, S. E. (2011). Persistence of soil organic matter as an ecosystem property. *Nature, 478*, 49-56.

Siddiky, R. K., Kohler, J., Cosme, M., and Rillig, M. C. (2012). Soil biota effects on soil structure: interactions between arbuscular mycorrhizal fungal mycelium and collembola. *Soil Biology and Biochemistry, 50*, 33-39.

Thakur, M. P., and Geisen, S. (2019). Trophic regulations of the soil microbiome. *Trends in microbiology*, *27*(9), 771-780.

Thornthwait, C. W. (1948). An approach toward a rational classification of climate. *Geographical Review, 38*(1), 55-94.

Van Oijen, M. (2008). Bayesian Calibration (BC) and Bayesian Model Comparison (BMC) of process-based models: Theory, implementation and guidelines.

Van Oijen, M., Rougier, J., and Smith, R. (2005). Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology, 25*(7), 915-927.

Vereecken, H., Schnepf, A., Hopmans, J. W., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M. H., Amelung, W., Aitkenhead, M., Allison, S. D., Assouline, S., Baveye, P., Berli, M., Brüggemann, N., Finke, P., Flury, M., Gaiser, T., Govers, G., Ghezzehei, T., Hallett, P., Hendricks Franssen, H. J., Heppell, J., Horn, R., Huisman, J. A., Jacques, D., Jonard, F., Kollet, S., Lafolie, F., Lamorski, K., Leitner, D., McBratney, A., Minasny, B., Montzka, C., Nowak, W., Pachepsky, Y., Padarian, J., Romano, N., Roth, K., Rothfuss, Y., Rowe, E. C., Schwen, A., Šimůnek, J., Tiktak, A., Van Dam, J., van der Zee, S. E. A. T. M., Vogel, H. J., Vrugt, J. A., Wöhling, T., and Young, I. M. (2016). Modeling soil processes: Review, key challenges, and new perspectives. *Vadose Zone Journal*, *15*(5).

von Lützow, M., Kögel-Knabner, I., Ludwig, B., Matzner, E., Flessa, H., Ekschmitt, K., Guggenberger, G., Marschner, B., and Kalbitz, K. (2008). Review Article Stabilization mechanisms of organic matter in four temperate soils: Development and application of a conceptual model. *Journal of Plant Nutrition and Soil Science, 171*(1), 111-124.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., and Bonan, G. B. (2014). Integrating microbial physiology and physio-chemical principles in soils with the MIcrobial-MIneral Carbon Stabilization (MIMICS) model. *Biogeosciences, 11*, 3899-3917.

Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G., and Bonan, G. B. (2015). Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development Discussions, 8*(2), 1789-1808.

Zhang, D., Hui, D., Luo, Y., and Zhou, G. (2008). Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. *Journal of Plant Ecology*, *1*(2), 85-93.

Zhao, B., Chen, J., Zhang, J., and Qin, S. (2010). Soil microbial biomass and activity response to repeated drying–rewetting cycles along a soil fertility gradient modified by long-term fertilization management practices. *Geoderma*, *160*(2), 218-224.

Table 1(on next page)

Initial input data.

Data from Brasschaat Scots pine forest (Belgium). Microbial C pool was estimated as hot water extractable C (HWC).

Variable	Unit	Value	Reference
Earthworm biomass	g C m^{-3}	200	Muys (1993)
pH		3.5	Janssens *et al*. (1999)
Sand	%	93	Janssens *et al*. (1999)
Initial SOM	g C m^{-3}	11470	Janssens *et al*. (1999)
Initial litter	g C m^{-3}	2680	Janssens *et al*. (1999)
Fine root biomass	g C m^{-3}	400	Janssens *et al*. (2002)
Fine root litter	g C m^{-3}	300	Janssens *et al*. (1999)
Fine root growth rate	g C m^{-3} $year^{-1}$	210	Janssens *et al*. (2002)
Annual litter fall	g C m^{-3} $year^{-1}$	400	Horemans *et al*. (2017)
Fine root turnover	g C m^{-3} $year^{-1}$	740	Based on Janssens *et al*. (2002)
C input to mycorrhiza	g C m^{-3} $year^{-1}$	197	Assumed based on Deckmyn *et al*. (2014)
Microbial C as HWC	g m^{-3}	1338.21	Gaublomme *et al*. (2006)

1

Table 2 (on next page)

Calibration data.

Data of C pools used for the model calibration. Biomasses of the nine food web functional groups: bacteria (B_b), fungi (B_f), mycorrhiza (B_{myc}), bacterivores (B_{bvores}), fungivores (B_{fvores}), detritivores (B_{det}), engineers (B_{eng}), herbivores (B_{hvores}) and predators (B_{pred}); and the other two C pools: litter and soil organic matter (SOM). Values were used once per year during calibration at days 180, 545, 910, 1275, 1640, 2005, 2370, 2735 and 3100.

C pool	Value (g C m^{-3})	Error (g C m^{-3})
$\mathbf{B_b}$	15.1	3.02
$\mathbf{B_f}$	15.1	3.02
$\mathbf{B_{myc}}$	160	32
$\mathbf{B_{bvores}}$	0.1	0.02
$\mathbf{B_{fvores}}$	0.8	0.16
$\mathbf{B_{det}}$	0.6	0.12
$\mathbf{B_{eng}}$	0.2	0.04
$\mathbf{B_{hvores}}$	0.2	0.04
$\mathbf{B_{pred}}$	0.4	0.04
Litter	2680	335
SOM	11470	1433.75

1

Table 3 (on next page)

Lower and upper bounds for the g_{max} prior probability distribution, for each one of the nine functional groups in the food web.

gmax	Lower bounds	Upper bounds
Bacteria	1	3
Fungi	0	3
Mycorrhiza	1	3
Bacterivores	0	2
Fungivores	0	2
Detritivores	0	0.5
Engineers	0	0.5
Herbivores	0	0.5
Predators	0	0.5

1

Table 4(on next page)

Averages ± standard deviation from the sample of 100 g_{max} vectors from the posterior distribution of the KEYLINK model calibrated for the Brasschaat Scots pine forest.

	Sample g_{max}
Bacteria	1.970 ± 0.424
Fungi	0.295 ± 0.134
Mycorrhiza	2.208 ± 0.302
Bacterivores	0.205 ± 0.098
Fungivores	0.095 ± 0.050
Detritivores	0.091 ± 0.050
Engineers	0.292 ± 0.062
Herbivores	0.028 ± 0.018
Predators	0.408 ± 0.060

1

Table 5 (on next page)

Effect of changes in input parameters on the average C pool (in g C m^{-3}) size over 10 years.

Averages from 100 runs of ten years with g_{max} parameter sets of the sample from the posterior distribution. The "basal" column has the results using reference input parameters (**Supplemental File S2**), and the other columns show the results with lower litter recalcitrance (rec 20%), lower input litter C:N ratio ($C:N_{lit}$ 40), higher pH (5.9), excluding predators (B_{pred} 0) and a higher clay content in the soil (clay 15%), respectively.

C pools	basal	rec 20%	CN_{lit} 40	pH 5.9	B_{pred} 0	clay 15%
Bacteria (g C/m^3)	5,98	6,94	6,35	5,90	2,67	4,75
Fungi (g C/m^3)	210,14	224,83	231,33	158,06	30,70	141,83
Mycorrhiza (g C/m^3)	39,83	40,20	40,03	38,70	29,75	39,27
Bacterivores (g C/m^3)	0,00	0,00	0,00	0,00	0,03	0,00
Fungivores (g C/m^3)	0,30	0,31	0,39	0,18	0,90	0,40
Detritivores (g C/m^3)	2,49	2,12	2,01	3,83	145,39	0,87
Engineers (g C/m^3)	0,04	0,04	0,04	0,51	1,54	0,05
Herbivores (g C/m^3)	0,12	0,12	0,13	0,02	5,48	0,16
Predators (g C/m^3)	3,33	2,86	2,87	6,41	0,00	1,36
Litter (g C/m^3)	3695,08	3262,71	3481,79	3728,85	2727,81	4245,54
SOM (g C/m^3)	10825,04	10941,01	10742,97	9347,23	3175,08	11655,70

1

2

3

Table 6(on next page)

Effect of changes in input parameters on the posterior distribution of C pool (in g C m^{-3}) size over 10 years.

Minimum and maximum values within 100 runs of ten years with g_{max} parameter sets of the sample from the posterior distribution. The "basal" columns have the results using reference input parameters (**Supplemental File S2**), and the following columns show the same changes from basal as in **Table 5**. For C pool notation see **Table 2**.

C pools	basal		rec 20%		CN_{lit} 40		pH 5.9		B_{pred} 0		clay 15%	
(g C/m^3)	min	max	min	max	min	max	min	max	min	max	min	max
B_b	9,61E-42	670,64	3,41E-43	676,40	8,48E-38	683,87	0,00	669,74	0,00	547,97	8,06E-43	788,31
B_f	2,63E-25	4660,93	3,87E-25	4587,16	3,74E-25	4677,78	0,00	4656,39	0,00	4038,09	2,56E-25	4763,21
B_{myc}	4,18	649,45	5,11	684,88	5,79	652,73	0,00	648,45	0,00	592,13	5,87	515,60
B_{bvores}	6,89E-110	0,10	4,91E-105	0,10	1,17E-108	0,10	0,00	0,10	0,00	44,07	1,13E-96	0,10
B_{fvores}	4,47E-84	194,08	5,82E-84	169,06	5,28E-83	204,50	0,00	141,68	0,00	351,94	1,27E-74	244,39
B_{det}	1,13E-67	1720,93	3,76E-67	1524,58	3,02E-68	1567,73	0,00	6898,50	0,00	5997,90	7,56E-79	431,85
B_{eng}	7,13E-178	61,50	6,54E-176	33,87	5,45E-172	74,46	0,00	396,62	0,00	88,62	7,18E-140	43,86
B_{hvores}	1,11E-68	49,65	5,23E-65	49,75	1,78E-65	49,64	3,78E-89	34,32	7,95E-22	85,00	1,12E-56	51,33
B_{pred}	4,72E-21	1616,28	4,73E-21	1409,80	4,72E-21	1466,09	0,00	5512,93	0,00	0,28	4,60E-21	452,07
Litter	6,0E+01	9056,63	42,89	8850,72	53,95	8996,91	60,08	8749,61	63,63	7409,28	75,00	9257,38
SOM	2,35E+3	15687,9	2356,64	16027,4	2350,89	15751,5	921,46	15687,4	1340,6	11589,	3919,25	15842,3

		7		9		6		4	6	4		9

1

Table 7 (on next page)

Effect of changes in input parameter on major output fluxes over 10 years.

The first three rows show bacterial, fungal and mycorrhiza respiration (R) fluxes (g C m^{-3}), respectively. The next three rows show the total turnover (g C m^{-3}) on an organic matter pool carried out by bacteria (Bact) or engineers (Eng) over 10 years. The penultimate row shows the fungi to bacteria ratio. And the last row is soil water content (SWC, l m^{-3}). Columns show average values and standard deviations from 100 runs of ten years from the sample of the posterior distribution, with a basal scenario using reference input parameters (**Supplemental File S2**), and the same changes from it as in **Table 5**.

	units	basal	rec 20%	$C:N_{lit}$ 40	pH 5.9	B_{pred} 0	clay 15%
R_{bact}	g C m^{-3}	250.2±222.8	293.6±227.2	262.3±225.7	248.9±222.4	119.3±149.9	204.5±225.9
R_{fun}	g C m^{-3}	3924.8±5043.6	4190.0±5035.6	4215.4±5166.1	3400.6±4776.3	963.6±2669.4	2638.5±4270.7
R_{myc}	g C m^{-3}	773.9±443.7	785.3±447.3	779.1±445.9	760.4±431.3	560.3±335.5	760.5±447.5
Bact SOM turnover	g C m^{-3}	995.6±758.2	1100.2±716.3	1036.5±757.7	985.4±752.6	438.6±524.2	780.2±741.7
Bact litter turnover	g C m^{-3}	258.6±264.1	367.7±328.6	280.2±276.2	257.6±263.9	119.6±170.6	223.5±279.7
Eng litter turnover	g C m^{-3}	1.5±5.3	1.3±4.5	1.3±4.4	43.8±104.4	97.3±72.1	2.0±5.8
B_{fungi}/B_{bact}	-	35.2	32.4	36.4	26.8	11.5	29.9
SWC	l m^{-3}	147.9±32.7	149.2±32.4	150.3±33.7	144.2±31.1	134.2±21.8	321.6±27.2

1

Figure 1

Simplified model scheme.

General structure of the KEYLINK model. Square boxes represent pools of organic matter. Wide double-line arrows, with a circle within the arrow, represent fluxes between pools (blue arrowheads show bidirectional fluxes). Isolated circles represent abiotic factors that affect model fluxes, and red narrow arrows connect each factor or pool with the model parts (at the arrowheads) that are regulated by them.

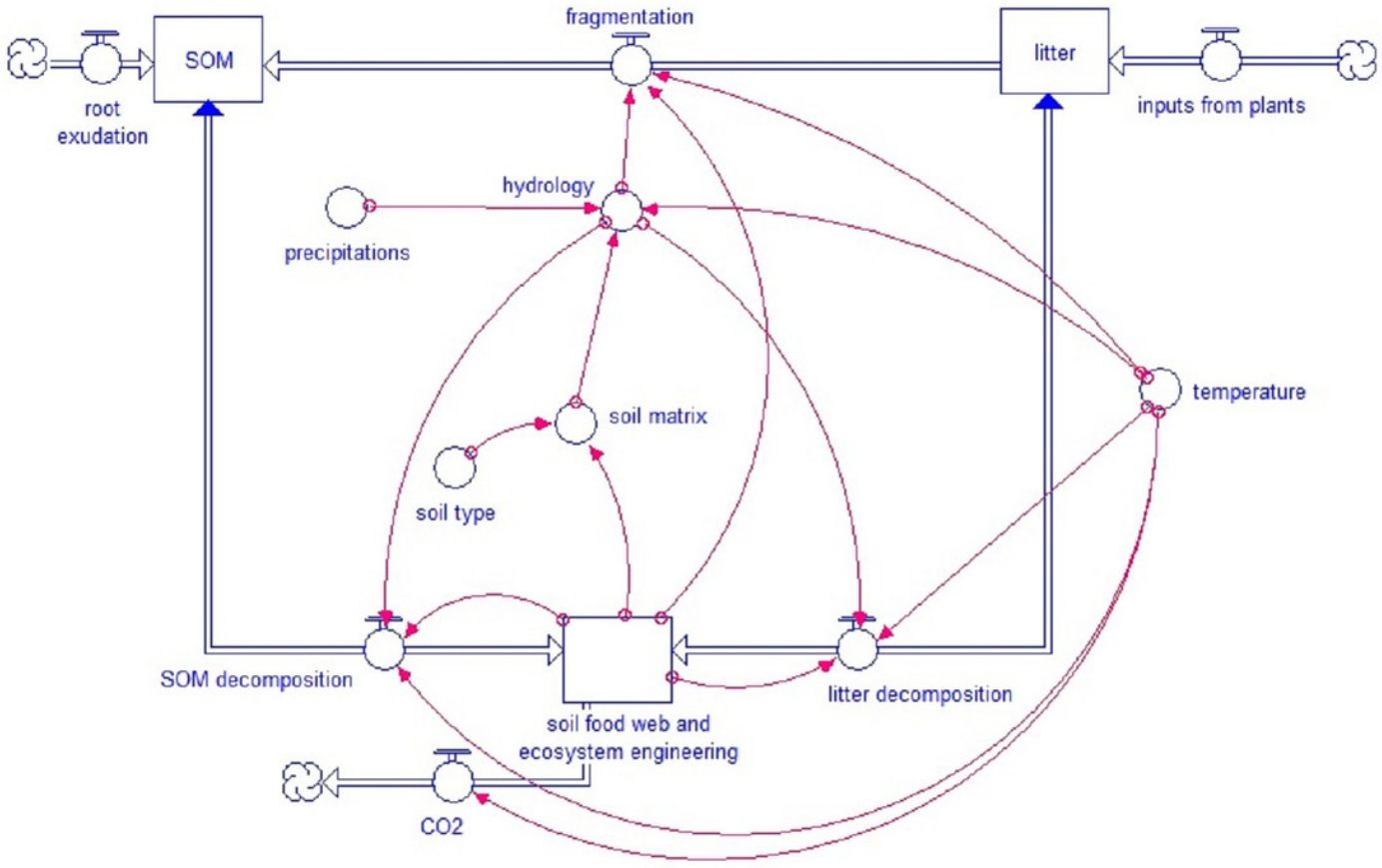

Figure 2

Pools and fluxes in the KEYLINK model.

Scheme of C pools with their interactions. All pools, soil, microorganisms and fauna (see **Table S1.2** in **Supplemental File S1**) are represented in the model in the same units (g C m^{-3}). The arrows represent carbon fluxes between the pools; each arrow is represented by a term in the model equations.

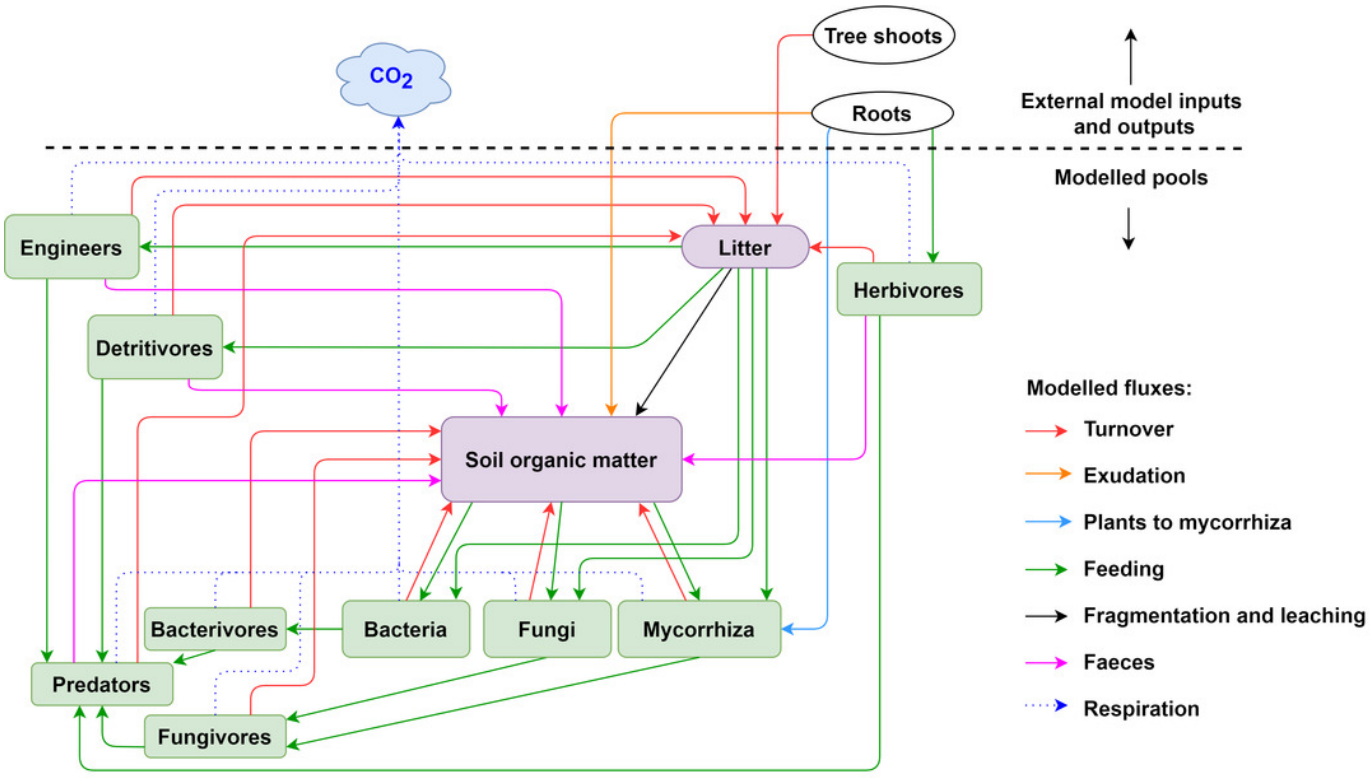

Figure 3

C pools daily biomass averages and standard deviations from the basal scenario

C pools (in g C m^{-3}) averages (black) and standard deviations (grey) among 100 simulations of ten years using the g_{max} sample from the basal simulation scenario. **A**) bacteria pool; **B**) non-mycorrhizal fungi; **C**) mycorrhizal fungi; **D**) microbivores feeding on bacteria; **E**) microbivores feeding on fungi; **F**) non-engineer detritivores; **G**) ecosystem engineers; **H**) herbivores; **I**) predators; **J**) plant litter; **K**) soil organic matter.

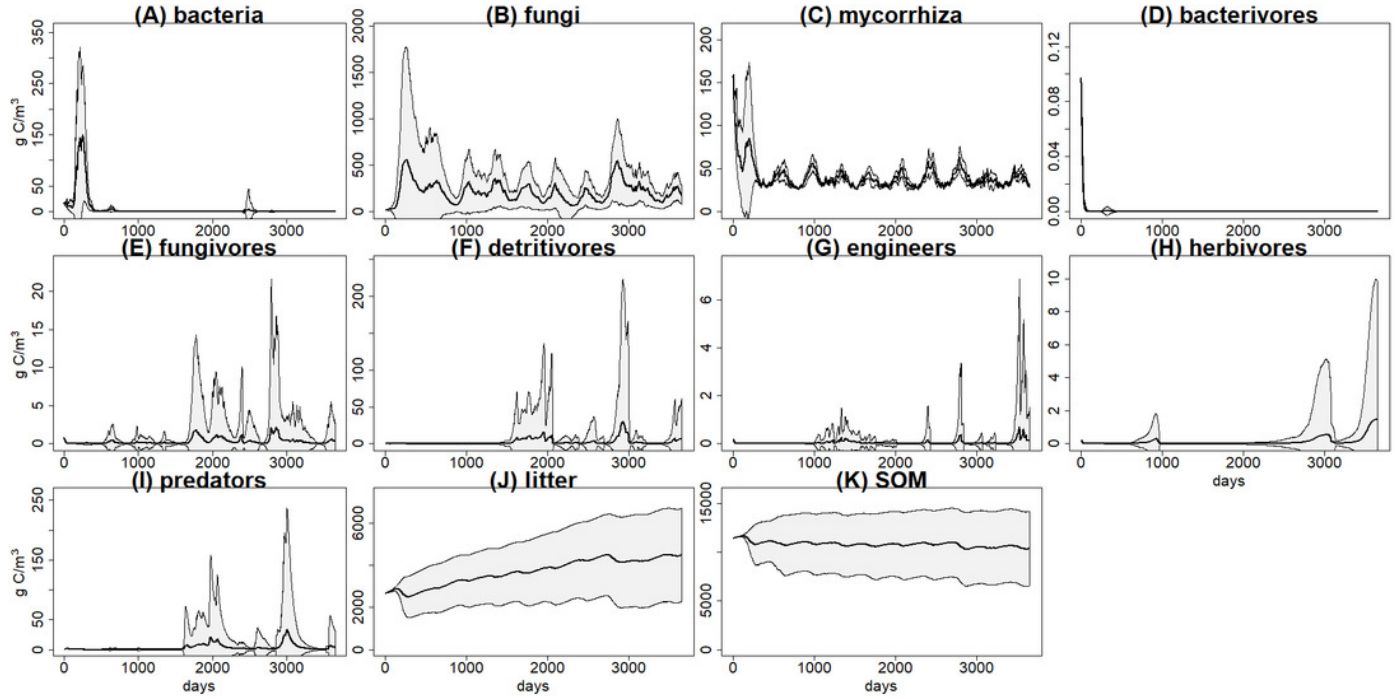

Figure 4

C pools daily biomass averages under different scenarios.

Averages of C pools (in g C m^{-3}) among 100 simulations of ten years using the g_{max} sample, with the basal simulation (black solid lines), and the alternative scenarios: higher clay content (red dashed lines), lower input litter C:N (yellow dotted lines), excluding predators (dark blue dotted dashed lines), higher soil pH (green long dash lines) and lower litter recalcitrance (light blue, lines with two different dashes). Pools are the same as in **Figure 3**, i.e. the nine food web functional groups, litter and soil organic matter (SOM).

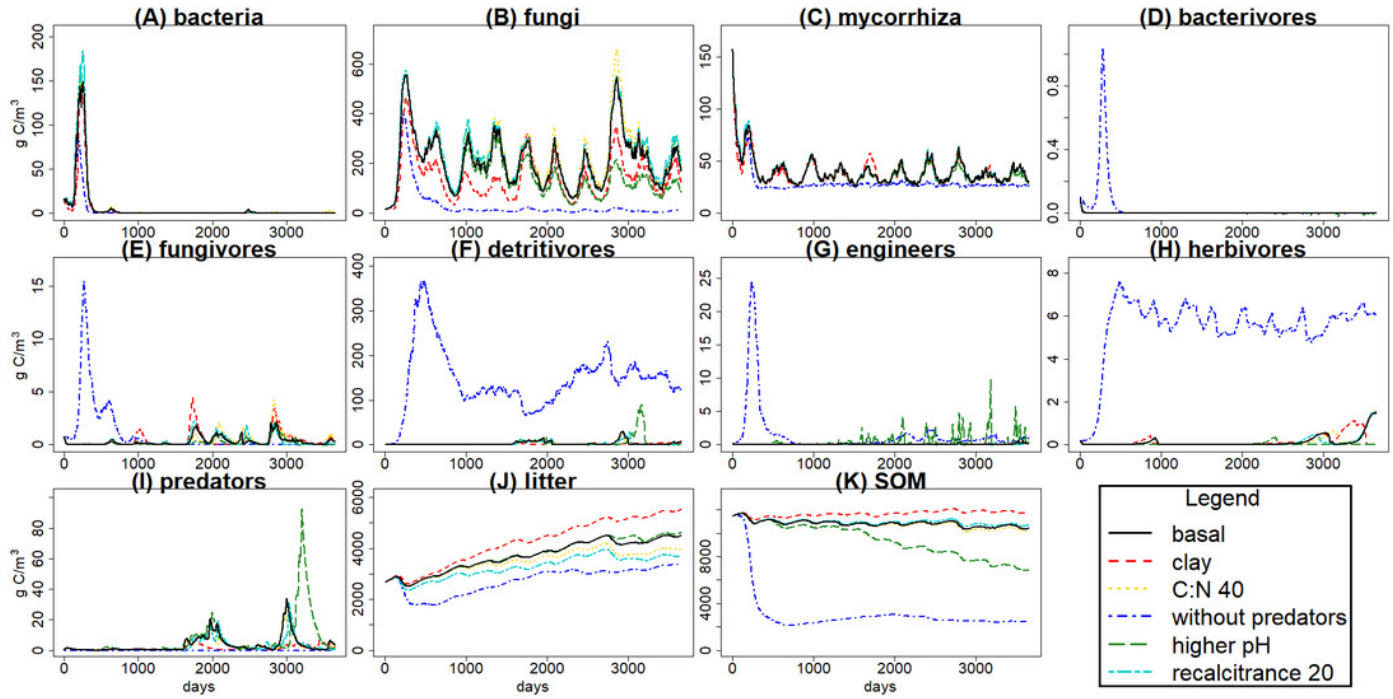

Figure 5

Daily volume averages of soil water content (SWC) and pore size classes in the soil matrix.

Means of volume (in l m^{-3}) among 100 simulations of ten years using the sample of g_{max} vectors, for the evaluation scenarios (see **Figure 4**). Graphs **A**-**D** for pore size classes, **E** for SWC. The innaccesible pore size class is not shown because it was not affected by changes in porosity.