# 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: omarf@mncn.csic.es

In the KEYLINK model, we integrate new knowledge on soil structure and its importance for hydrology and soil organic matter (SOM) stabilization, 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, one of the first and most ambitious attempts to integrate soil functional diversity and food webs in predictions of soil carbon (C) and soil water balances. In addition, this mechanistic process-based model can be coupled to other ecosystem models and improve their predictions as an alternative to the widely used more chemically based models. 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 C 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, in contrast with broadly used first order kinetic models, 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.

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

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

9 [1]Department of Biogeography and Global Change, National Museum of Natural Sciences -
10 Spanish National Research Council (MNCN-CSIC), Madrid, Spain

11 [2]Centre of Excellence PLECO, Department of Biology, Universiteit Antwerpen, Antwerp,
12 Belgium

13 [3]BC3-Basque Centre for Climate Change, Leioa, Spain

14 [4] IKERBASQUE - Basque Foundation for Science, Bilbao, Spain

15 [5]Earth and Life Institute, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

16 [6]Agrosphere Institute (IBG-3), Forschungszentrum Jülich GmbH, Jülich, Germany

17 [7]Laboratory of Soil Zoology, A.N. Severtsov Institute of Ecology and Evolution, Russian
18 Academy of Sciences, Moscow, Russia

19 [8]Forest Research, Farnham, United Kingdom

20 [9]Department of Environment and Climate, Research Institute for Nature and Forest, Brussels,
21 Belgium

22 [10]ARAID, Pyrenean Institute of Ecology - Spanish National Research Council (IPE-CSIC), Jaca,
23 Spain

25 **Corresponding Author:**

26 Omar Flores[1, 2]

27 Universiteitsplein 1, Wilrijk, 2610, Belgium

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

## 30 Abstract

31 In the KEYLINK model, we integrate new knowledge on soil structure and its importance for
32 hydrology and soil organic matter (SOM) stabilization, with the existing concepts on SOM
33 pools, and elements from food web models, i.e. those from direct trophic interactions among soil

34  organisms. KEYLINK is, therefore, one of the first and most ambitious attempts to integrate soil
35  functional diversity and food webs in predictions of soil carbon (C) and soil water balances. In
36  addition, this mechanistic process-based model can be coupled to other ecosystem models and
37  improve their predictions as an alternative to the widely used more chemically based models. We
38  present a selection of equations that can be used for most models as well as basic parameter
39  intervals for, e.g., key pools, functional groups' biomasses and growth rates. Parameter
40  distributions can be determined with Bayesian calibration, and here an example is presented for
41  food web growth rate parameters for a pine forest in Belgium. We show how these added
42  equations can improve the functioning of the model in describing known phenomena. For this,
43  five test cases are given as simulation examples: changing the input litter quality (recalcitrance
44  and C to nitrogen ratio), excluding predators, increasing pH and changing initial soil porosity.
45  These results overall show how KEYLINK is able to simulate the known effects of these
46  parameters and can simulate the linked effects of biopore formation, hydrology and aggregation
47  on soil functioning. Furthermore, the results show an important trophic cascade effect of
48  predation on the complete C cycle with repercussions on the soil structure as ecosystem
49  engineers are predated, and on SOM turnover when predation on fungivore and bacterivore
50  populations are reduced. In summary, in contrast with broadly used first order kinetic models,
51  KEYLINK shows how soil functional diversity and trophic organization and their role in C and
52  water cycling in soils should be considered in order to improve our predictions on C
53  sequestration and C emissions from soils.

54

## Introduction

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

69  The insights concerning the role of the microbial biomass in C turnover has been introduced in
70  models such as MIcrobial-MIneral Carbon Stabilization (MIMICS) model (Wieder et al., 2014;
71  Wieder et al., 2015) and Litter Decomposition and Leaching (LIDEL) model (Campbell et al.,
72  2016). However, soil fauna and especially ecosystem engineers, sensu Jones et al. (1994), have
73  also been shown to play a key role in determining C and nutrient turnover and hydrology of soils
74  through their impact on aggregation, pore formation and bioturbation as well as their direct
75  contribution to litter and SOM turnover (Filser et al., 2016; Lavelle et al., 2016). Several authors
76  have highlighted the need to include soil fauna contributions to SOM dynamics into soil
77  modelling (see review by Vereecken et al., 2016). This information has been used in detailed and

78    small-scale soil models (Chertov *et al*., 2017; Geisen *et al*., 2019), but is not incorporated into
79    larger-scale ecosystem models. The main difficulty is the lack of data concerning the soil, either
80    physical, chemical or biological, and the different methods used, making parameterization of any
81    model unsure. The goal of the KEYLINK model is to consider the soil including the main
82    mechanisms concerning the effects of soil biota on litter and SOM transformations and
83    hydrology through structural modifications, without increasing the number of parameters beyond
84    what is currently available on most well-measured ecosystems (G Deckmyn *et al*., 2020,
85    unpublished data). We show how this model is parameterized for a forest stand where soil fauna
86    was never studied in detail, but many other soil and stand characteristics are well established.

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

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

101

## Methodology

### Structural effects

104   Pore size distribution determines accessibility for trophic interactions of soil fauna and soil
105   microorganisms (**Fig. 2**), both by size and by aeration and $H_2O$; soil fauna changes pore size
106   distribution and produces cracks and fissures in the soil. In the model, pore size distribution is
107   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 all the following pore categories) 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. Macroporosity is hardly measurable with typical lab measurements or the retention curve but visual assessment is possible.

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 our 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$^{-1}$). The daily loss of water by evapotranspiration is calculated using an equation for potential evapotranspiration based on Thornthwait (1948). 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 $P_{net}$ .

157 $I + P_{runoff} = P_{net}$ (4)

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

165 $I_{maxPor} = PV_{macro} + I_{maxMat(n+1)}$ (5)

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

167 $P_{runoff} = P_{net} - (I_{maxPor} + I_{maxMat})$ (6)

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

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

175 $SW_n = SW_n + \min(PV_n - SW_n, I_{maxmat}(1 - SA_{macro}), Pnet(1 - SA_{macro}))$ (7)

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

178 $D_n = P_{net} SA_{macro} + P_{net} -$

179 $\min(PV_n - SW_n, I_{maxmat}(1 - SA_{macro}), P_{net}(1 - SA_{macro}))$ (8)

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

182

183 C flow

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

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

199 $G = \Sigma\left(g_{max}\left(\dfrac{Sf_a}{K_s + S}\right)\right)B$ (9)

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

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

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

223 $\dfrac{dB}{dt} = G - R - D_t - P_d$ (10)

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

230 $R = rB$  (11)

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

232 $D_t = dB$  (12)

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

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

236

237  Effect of H$_2$O

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

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

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

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

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

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

253

254  Simulating the variability in g$_{max}$

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

265

266 Simulating the effect of temperature (T)

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

271
$$m_T = \begin{cases} 0, & T < T_{min} \text{ or } T > T_{max} \\ Q^{(T - T_{opt})/10}, & T_{min} < \qquad T < T_{opt} \\ 1, & T_{opt} < \qquad T < T_{max} \end{cases} \quad (17)$$

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

276

277 Effect of pH on growth

278 A good example of an optional effect is the effect of pH: for a system close to a threshold,
279 simulating pH can be very important, assuming a good knowledge of the system. But for well-
280 buffered systems, it is an unnecessary increase in complexity. $g_{max}$ decreases at low pH for
281 bacteria but increases for fungi. For this example, we put the thresholds at 8 for fungi and 3 for
282 bacteria inducing a 10 fold reduction in $g_{max}$ for a change in pH of 1.

283 $m_{pH} = 1/((pH - 8)10)$     for fungi if $pH > 8$    (18)

284 $m_{pH} = 1/((3 - pH)10)$     for bacteria if $pH < 3$      (19)

285 In any other case for bacteria or fungi, $m_{pH} = 1$. For engineer saprotrophs, their optimal $g_{max}$
286 changes (becoming $g_{maxEng}$) with pH according to the following equation:

287
$$g_{maxEng} = \begin{cases} 0, & \text{if } pH < 3 \\ \left(\frac{g_{max}}{2}\right)(pH - 3), & \text{if } 3 \leq pH < 5 \\ g_{max}, & \text{if } pH \geq 5 \end{cases} \quad (20)$$

288

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

290 Overall consumption of an organism that can consume different pools is computed by simply
291 adding them up. However, litter is not necessarily as 'palatable' depending on C:N ratio, if not
292 enough N then it is needed to consume more, and recalcitrance, if low in energy it is needed to
293 consume more, through modifiers $m_{C:N}$ and $m_{rec}$. This is simulated by changing $g_{max}$. The
294 equation for $m_{rec}$ is not necessary and only important if enough data on litter quality is available
295 or the users are interested into looking into the effects of changes in litter quality. The litter pool

296    can be consumed by both bacteria and fungi, and of course also detritivores. Depending on the
297    C:N ratio, the competition between these two is different; this is simulated by the $g_{max}$ of the
298    bacteria being more variable with C:N ratio. The sensitivity is described by the parameter $p_{mC:N}$,
299    between 0 and 1.

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

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

302

303    For litter recalcitrance ($Rec_{lit}$), a linear equation instead of a power is chosen so that decay of the
304    recalcitrant litter is 0 if $p_{mRec} = 1$ and is unaffected if $p_{mRec} = 0$.

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

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

307

308    Adding up all these modifying effects on $g_{max}$

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

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

313

314    Closing the C budget

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

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

324

325    Closing the budget of recalcitrance and N

326    KEYLINK can be run with or without a detailed N model, but if a detailed N model is included,
327    it is important to close the N budget; if not the following equations need not be used. Including a
328    closed recalcitrance budget if N is not included seems an unlikely choice since N is more

329   important than recalcitrance as a driver. Respiration reduces the C of the biomass pool of the
330   organism. The associated N goes into excretion for larger fauna; for microbes, it is mineralised to
331   ammonium (i.e. plant available mineral N) unless it is necessary for growth, if the C:N ratio of
332   the SOM is higher than that of the biota after deducting respiration. For microbes growth on a
333   source not containing enough N is possible but the ratio respiration to growth will increase to
334   close the budget.

335 $\quad g_{max} = r + \dfrac{C:N_{SOM}}{C:N_{bact}} r \quad (27)$

336   When there is no N simulated, this equation will stop almost all bacterial growth, because litter
337   has a much higher C:N ratio. In this case, we recommend to replace equation (27) by calibrating
338   the $m_{C:N}$. The faeces recalcitrance and C:N ratio are calculated from the difference between
339   consumer and consumed source for the larger fauna. To close the N cycle it is important that the
340   consumer cannot grow if there is not enough N to build its tissues: faeces can have a lower or
341   higher N content depending on the fraction to faeces ($f_{faec}$), which increases if N is limiting. This
342   is only relevant for detritivores and engineers, because the others eat each other at constant C:N
343   ratios. Under N limitation, N consumed plus available from respiration surplus needs to equal N
344   used for growth.

345 $\quad \dfrac{r + (1 + f_{faecC:N})g_{max}}{C:N_{SOM}} = \dfrac{g_{max}}{C:N_{eng}} \quad (28)$

346 $\quad f_{faecC:N} = \dfrac{C:N_{SOM}}{C:N_{eng}} - \dfrac{r}{g_{max}} - 1 \quad (29)$

347 $\quad f_{faecEff} = max(f_{faec}, f_{faecC:N}) \quad (30)$

348   For recalcitrance of SOM, if enough data are available to include this, assume faeces are twice as
349   recalcitrant as the consumed pool. Recalcitrance of each faunal group is an input constant.

350

351   Calculations regarding engineers

352   Soil changes made by engineer species depend on their body width, but in the model this is
353   simplified using initial parameters for engineers' effects that must be chosen based on an average
354   width (see **Table S2.5** in **Supplemental File S2**); the model then simulates their daily effects
355   using their biomass. Bioturbation is a function of engineer biomass ($B_{eng}$, g C m$^{-3}$), which
356   calculates organic matter moving to deeper layers: litter moving (g $C_{lit}$ g $C_{eng}^{-1}$ day$^{-1}$) from litter
357   layer to end of burrow, and SOM moving by mixing of soil between layers (g $C_{SOM}$ g $C_{eng}^{-1}$ day$^{-1}$
358   $^{1}$). In this first version of the model, with only one soil layer, bioturbation works as a C output
359   flow, but in future versions with more layers it could be upgraded to C flows between them.

360   Burrow volume ($PV_B$, l m$^{-3}$) is a function of engineer biomass and the ratio of pore volume to
361   engineer biomass ($VE_{ratio}$, l g $C_{eng}^{-1}$) but towards a maximum ($PV_{Bmax}$, l m$^{-3}$):

362 $\quad PV_B = min(PV_{Bmax}, VE_{ratio}B_{eng}) \quad (31)$

363   On the other hand, burrow turnover happens at a constant rate, with average burrow lifespan of
364   10 years; porosity decreases and burrows become mesopores. This could be improved in future
365   versions including perturbations as the possible effect of heavy rain.

366

367  Porosity calculations

368  The pore volume is distributed in five classes by pore size. Initial pore size distribution is given
369  or measured as the total pore volume (PV, $l\,m^{-3}$) in each class. The link between aggregation and
370  porosity is hard to quantify. Regelink et al. (2015a) showed for different soils that overall soil
371  porosity is the sum of the textural porosity determined by the proportion of clay, sand and silt
372  fractions and aggregation porosity. They conclude that micropores, which they define <9 μm, are
373  mainly situated within the aggregates, while mesopores are situated between dry-sieved
374  aggregates. While Regelink *et al.* (2015a) have shown that total micro and mesoporosity (<1000
375  μm) increases with total aggregate content, Grosbellet *et al.* (2011) have provided evidence that
376  pores in the range 30 – 300 μm decrease with aggregation. Despite of the generally lower ranges
377  for mesopores (9 – 1000 μm) described for soil physics (Lal and Shukla, 2004; Regelink *et al.*,
378  2015a), we want mesopores to be physically accessible to mesofauna body size (ca. 100 – 2000
379  μm), so we consider that mesopores ranging 30 – 1500 μm are a reasonable compromise. Based
380  on that, we decided to hypothesize that aggregation increases bacterial and micro- porosity while
381  decreasing mesoporosity. However, we want to emphasize that further experimental studies are
382  needed to establish robust relationships between aggregation and pore size distribution.

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

386  • Textural  porosity ($PV_{text}$): measured or calculated from % clay and sand.
387  • Additional aggregation porosity ($PV_{Ag}$): all porosity in surplus of textural, can be
388    estimated, for example from PTF (PedoTransfer Function) or calculated empirically from
389    SOM and fungal biomass, i.e. mycorrhiza and other fungi, max 2% porosity extra
390    (equation 33). Aggregation (Ag) is the fraction (0–1) of the SOM aggregated calculated
391    as (based on the data from Malamoud *et al.*, 2009):

392  $$Ag \;=\; min(1, \frac{c(B_{fung} + B_{myc})}{B_{SOM}}) \qquad (32)$$

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

394  $$PV_{Ag} \;=\; k\,Ag\,B_{SOM} \quad (33)$$

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

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

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

405    $PV_{tot} = PV_{text} + PV_{Ag} + f_{PV}PV_B$      (34)

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

407    The volume of micropores ($PV_{micro}$) and bacterial pores ($PV_{bact}$) increases with increasing
408    aggregation. Apart from creating additional porosity depending on the total amount of
409    aggregated SOM (eq. 33), aggregation also increases the relative micro- and bacterial pore
410    volume at the expense of (textural) mesoporosity ($PV_{meso}$), therefore not increasing total
411    porosity. This effect is controlled by available pore space between mineral particles (i.e. textural
412    mesoporosity) and we assume that half of this mesoporosity can be affected by aggregation. In
413    both cases, we assume that the increase in porosity due to aggregation is divided equally among
414    micropores and bacterial pores. The pore volume in different size classes is calculated as:

415    $PV_{macro} = PV_{textmacro} + PV_B$      (35)

416    $PV_{meso} = PV_{textmeso} - (1 - f_{PV})PV_B - \frac{Ag}{2}PV_{textmeso}$      (36)

417    $PV_{micro} = PV_{textmicro} + k\frac{Ag}{2}B_{SOM} + \frac{Ag}{4}PV_{textmeso}$      (37)

418    $PV_{bact} = PV_{textbact} + k\frac{Ag}{2}B_{SOM} + \frac{Ag}{4}PV_{textmeso}$    (38)

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

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

422

423    Leaching

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

429    DOM can be simulated in relation to the $CO_2$ released, high 'activity' in the soil, as total
430    respiration ($R_{tot}$, g C m$^{-3}$ day$^{-1}$) by the fraction of respiration from DOM ($f_{DOM}$), similar to the
431    concepts used in the LIDEL model, in addition to the directly exuded DOM ($C_{Exud}$, g C m$^{-3}$ day$^{-}$
432    $^1$). Assuming a short half-life of DOM and semi-empirically, because daily concentration is not
433    'equal' to daily production ($DOM_p$, g C m$^{-3}$ day$^{-1}$) but is linearly related to the daily production,
434    we consider:

435    $DOM_p = C_{Exud} + f_{DOM}R_{tot}$      (39)

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

440    $DOM_{ad} = K_D\ DOM$          (40)

441 In addition depending on the minerals and pH, more or less DOM and mineral N will be
442 dissolved. More clay (fClay fraction) means less mobile DOM, and lower pH is also a cause of
443 less mobile DOM.

444    $K_D = a_{KD} - b_{KD}pH + c_{KD}\ fClay$          (41)

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

447

448 Calculation order

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

451      a)   Calculate the pore size fractions in 5 classes and the associated pore surface areas
452      b)   Calculate the water volume of the relevant pore size
453      c)   Use the precipitation leaching to calculate DOM leaching
454      d)   Calculate for each biota group the accessibility of each of the pools it consumes
455      e)   Calculate the $g_{max}$ depending on temperature, $H_2O$, C:N, pH and recalcitrance
456      f)   Solve the 12 differential equations for increase/decrease of all C pools
457      g)   Update all C pools
458      h)   Calculate the new C:N and recalcitrance of each pool
459      i)   Calculate engineering effect
460          a.   Update macropores
461          b.   Update SOM from bioturbation
462      j)   Calculate other changes in pore size distribution from weather or management

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

469

470 Model coding and output

471 KEYLINK consists of a relatively limited, freely downloadable Python code (available at:
472 https://github.com/Plant-Root-Soil-Interactions-Modelling/KEYLINK) that is very easy to
473 modify and calibrate towards specific questions or ecosystems, and to link to existing ecosystem
474 models. Each of the modifiers on growth, i.e. temperature, pH, $H_2O$, recalcitrance and C:N, as
475 well as the primal shape of the growth equations can be adapted. The inputs in the current
476 version are read from data-files but are easy to link to a mechanistic model. The output of the
477 current version consist of all daily C pools as well as the main C fluxes. KEYLINK is also
478 available as a stand-alone executable model, allowing it to be called from models in other
479 languages. A single run of ten years could take less than one minute (depending on computing

480    power). We advise using at least a hundred runs to reach a more representative result; this takes
481    approximately one hour from a dataset of a hundred input parameter sets. In this version, the
482    average results over the hundred runs are calculated but also all daily outputs of each run are
483    saved.

484

485    Model parameterization

486    The first version of KEYLINK model has been parameterized for a Scots pine forest stand
487    situated in Brasschaat, in the Campine region in Belgium (51°18' N and 4°31' E). The soil is
488    sandy but with high ground water table so trees are generally not water-limited, but the topsoil is
489    often dry. The soil is acidic (pH 3.5). The trees were planted around 1930 and formed a rather
490    sparse vegetation in 1999, with leaf area index (LAI) ranging from 2.1 to 2.4.

491    For this model run, we used the following input data from the stand (**Table 1**). In this case, we
492    did not use measured or modeled growing trees but constant input of aboveground and
493    belowground litter to show how the KEYLINK model works by itself. Brasschaat forest data
494    concerning the top 90 cm of soil was analyzed in 1999 by Janssens *et al*. Earthworm biomass is
495    extremely low due to the low pH, it was not measured since 1993 by Muys, but these data are
496    used since there is no reason to expect there was a marked change.

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

504    Model calibration

505    Once the model is parameterized for an ecosystem, the next step is to optimize that model,
506    calibrating the fit of its simulations to the ecosystem data. The optimization included in the
507    KEYLINK model follows a Bayesian procedure as described by Van Oijen (2008). The Bayesian
508    method allows the determination of model parameters and their uncertainties by combining (1)
509    prior information about parameter values and uncertainty and (2) experimental observations of
510    output variables. The prior parameter information can be obtained directly from measurements or
511    it can be derived from the literature. We calibrated the model applying the Bayes' Theorem that
512    in a simplified form can be written as:

513    $$p(\theta \mid D) = c \cdot p(D \mid \theta) p(\theta) \qquad (42)$$

514    where $p(\theta \mid D)$ is the posterior distribution of the parameter value θ, c a constant ($1/p(D)$),
515    $p(D \mid \theta)$ is the likelihood function for θ and the factor p(θ) the prior distribution for θ (Van
516    Oijen *et al*., 2005). The data likelihood function is determined by the probability errors in
517    observations. We assumed that errors are uncorrelated and normally distributed with zero mean
518    (Van Oijen, Rougier and Smith, 2005). To avoid rounding errors, the logarithm was determined
519    as follows:

$$\log L_i = \sum_{j=1}^{n} \left( -\frac{1}{2} \left( \frac{D_j - f(\theta_j)}{\sigma_j} \right)^2 - \frac{1}{2} \log(2\pi) - \log(\sigma_j) \right)$$

520

(43)

521  Where $D_j$ is the observed data in sampling year j, $f(\theta_j)$ is the simulated value, and $\sigma_j$ the
522  standard deviation of the model error. Practically, the posterior parameter distribution was
523  estimated in a non-analytical way following Van Oijen, Rougier and Smith (2005). Model
524  calibration was run more than 20000 times (for each version, the first and the second adapted to
525  drylands) with different parameter settings sampled from the prior parameter distribution, using
526  the version of Markov Chain Monte Carlo (MCMC) known as the Metropolis-Hastings random
527  walk with reflection algorithm (Christian and Casella, 1999; Van Oijen, 2008).

528  The goal is to walk through parameter space (prior distribution) in such a way that the collection
529  of visited points forms a sample from of the calibrated parameter values (posterior distribution).
530  This Bayesian calibration scheme generates a chain of accepted parameter values and
531  corresponding model output.

532  A pragmatic assumption is that the starting values of the C pools (including the soil fauna initial
533  biomass) are at steady state for a given data (most often spring or summer). The simplest
534  calibration of any ecosystem can be done by assuming these 11 carbon pools (litter, SOM and
535  the 9 functional groups in food web) need to be stable over the simulated years, e.g. for 9 years
536  that gives us 99 data points by taking the same value for each C pool every year (**Table 2**). Initial
537  litter, SOM and biomasses of bacteria, fungi and engineers were taken from the references cited
538  in **Table 1**. For other C pools, data were estimated using measured data for previous C pools and
539  similar proportions between C pools as in the Swedish pine forest in Persson *et al.* (1980);
540  predator biomass was assumed to be the 20% of all biomass in their consumed C pools. Errors
541  were assumed as a percentage of biomass, 10% for predators, 12.5% for litter and SOM, and
542  20% for the rest C pools.

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

548

549  Input parameters of species

550  The KEYLINK model framework is conceptualized as an adaptable framework. Each user needs
551  to determine for their specific site and questions the main drivers and pools required. Depending
552  on the dataset, it is in general better to use less pools and equations if sparse data are available.
553  Moreover, KEYLINK is not a soil fauna model and was not designed to simulate specific soil
554  fauna species in detail. The soil fauna groups used consist of a wide range of species, for which
555  average data are used. For a description of the species categories, we refer to the review on the
556  KEYLINK concepts (G Deckmyn *et al.*, 2020, unpublished data).

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

567

568     Calibration for Brasschaat pine forest

569     We show here the results from a calibration towards data measured and assumed, using
570     proportions between fauna groups in Persson *et al*. (1980), in the Brasschaat Scots pine stand in
571     Belgium. This forest stand is relatively well described in many publications concerning the trees
572     and the total ecosystem fluxes, but less concerning the soil and very little was measured on soil
573     fauna. We use this forest as an example of how the KEYLINK model can be used to improve our
574     understanding of the system even when detailed soil faunal data are limited.

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

584     In our case, no measurements were available and information in the literature was scant.
585     Therefore, we deliberately defined wide ranges for the prior values of each parameter to cover all
586     the possible values found in the literature (Chuine, 2000; Linkosalo, Lappalainen and Hari,
587     2008). For species for which no prior parameter information was available, we assumed
588     parameter values equal to the mean value of the range. The initial uncertainty of each parameter
589     is quantified in terms of a prior probability distribution with lower and upper bounds. Because of
590     lack of detailed knowledge, we assumed the distribution as uniform and non-correlated.

591     The $g_{max}$ values were optimized using the prior range for $g_{max}$ (**Table 3**). The data used to
592     calibrate against were chosen to give a 'standard' procedure, so limited to biomass of the
593     different C pools. Including all available data s.a. soil respiration, soil humidity could improve
594     the run for Brasschaat, but would not be a representative run for the model. Other parameter
595     settings, e.g. sensitivity to C:N and recalcitrance, were based on model runs of the Brasschaat
596     site by Deckmyn *et al*. (2011).

597     We ran the model for the time period 1999-2008, because this was the period in which the forest
598     was still clearly dominated by Scots pine; since then a transition to more deciduous trees has

599    been taking place. We calibrated towards stable C pools over the ten years for all C pools, with
600    an allowed error margin of 20% for all faunal pools, except 10% for predators, and 12.5% for
601    litter and SOM. Daily climate data (temperature and precipitation) were used. The full range of
602    input data can be found in **Supplemental File S2**, except climate data, which can be downloaded
603    with the model. Choosing to calibrate towards one or more pools can yield different results, and
604    it depends on the end-user's goal which calibration is preferred.

605

606    Model evaluation

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

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

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

619    In each one of the five alternative scenarios, input parameters were the same than in the basal
620    scenario, except for the parameter changed to generate the new scenario (see **Supplemental File
621    S2**). All the six scenarios were run 100 times using the LHS, as mentioned before, consisting
622    each run in a simulation of 10 years at a daily time-step (3653 days). Then, averages of biomass
623    were calculated for each C pool among the 100 simulations of 10 years, for each scenario,
624    comparing the effects of disturbances on average values.

625

626    **Results**

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

630    The optimization clearly showed the strong link between the different groups of soil biota, e.g. a
631    high $g_{max}$ for bacteria was coupled to a high $g_{max}$ for bacterivores. Running the model 100 times
632    using the sample of the $g_{max}$ values resulted in predictions with a quite wide range. The
633    alternative five scenarios compared to the basal one can show very different results concerning
634    specific C pools (**Table 5**).

635    Mycorrhiza, herbivores and detritivores are relatively uncoupled, though influenced by
636    predators, and follow the yearly climate curves. The bacterial and fungal biomasses are very
637    strongly linked. The high $g_{max}$ of bacteria allows steep peaks, which are generally followed by
638    peaks in bacterivore biomass. As we used constant litter input into the soil and used a calculated

639 constant fraction of potential evapotranspiration as water uptake from the soil, it cannot be
640 expected that these results follow the normal annual trends in fluctuations of those fluxes. This
641 can, at least partially, explain the relatively low bacterial biomass found in our results since the
642 bacteria would profit most from a rapidly changing environment. For more realistic simulations
643 the model can be coupled with other models that give that information as outputs, or with
644 measured datasets.

645 All C pools tend to reach stability after the first years, suggesting the model is well-balanced;
646 however, stability values seem to be more sensitive to changes in $g_{max}$ parameters for some pools
647 (e.g. SOM). The set-up of the model, where we only calibrate the faunal $g_{max}$, does not allow
648 calibration towards different ratio of litter and SOM decay. This depends on the uncalibrated
649 parameter fragmentation, the sensitivity to recalcitrance, but also the temperature used for the
650 litter and SOM. Here we used the same temperature while in reality, since we use data from the
651 top 1 m of soil; a lower temperature for SOM compared to surface litter would be more realistic.

652 An overview of all C pools under the different simulation scenarios (**Fig. 3**) shows how changing
653 one input parameter at a time influences the results. It must be born in mind that KEYLINK was
654 run as a stand-alone model; linking it to a model or more detailed data of the aboveground
655 ecosystem would greatly influence the results, but would not allow clear interpretation of the
656 model functioning due to feedbacks. To further elucidate these effects and to show some of the
657 potential outputs the model can give, we show a few of the most interesting fluxes (**Table 6**).

658 Increasing clay content resulted in an increase in water content (**Fig. 4**) and a decrease in SOM
659 decay while fungal/bacterial ratio decreased.

660

## Discussion

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

668 The run without predators showed the most interesting results because the interactions between
669 the different food web parts are apparent. Exclusion of predators, setting the starting biomass of
670 that pool at 0, showed how the model tracks its crucial role in the ecosystem (**Fig. 3**). Predators
671 produce a top-down trophic cascade on the food web, e.g. on herbivores and roots. Microbial
672 decay is reduced as fungi and bacteria are consumed by the increased populations of bacterivores
673 and fungivores. Despite of the decrease in bacteria and fungi, SOM and litter were also lower
674 without predators. This could be explained by an increase in engineer populations with
675 implications also on the soil matrix, as we discuss in the next section. Overall the model
676 successfully tracked soil food web dynamics and also their interactions with soil porosity. The
677 effect of larger soil predators (e.g. Araneae, Carabidae, Formicidae) slowing down SOM
678 decomposition and enhancing its stabilization has been previously found in experiments (Kajak,
679 1995), as well as mycorrhiza effect on porosity by making aggregates (Siddiky *et al*., 2012).
680 Therefore, KEYLINK model  seems to fit with the expected food web and C dynamics, and

681 could serve to improve the biogeochemical cycles modelling, as is needed for larger scale
682 predictions (Grandy *et al.*, 2016), by coupling it with other ecosystem models.

683 Hydrology is influenced by aggregation and by macropore formation by ecosystem engineers.
684 The increased macroporosity increases infiltration rate with reduced water-logging and runoff.
685 Predators have a clear indirect effect on soil porosity by consuming engineer species, and also
686 microbivore species, which leads to changes in soil hydrology (**Fig. 4**). Variations in bacterial
687 pore and micropore volume are positively correlated, while mesopore variations are negatively
688 correlated with both; the higher volume in mesopores, the lower in the two other classes, and the
689 faster the water drains from the soil layer. That is what we can expect to happen in real soils, so
690 the model seems to simulate appropriately those dynamics. The increase in macro and
691 mesoporosity volumes without predators, so with higher engineers, resulted in a decrease of soil
692 water content of 9.26 % (increasing the pore aeration), and under those conditions the
693 availability of SOM and litter for bacteria and fungi could be increased, explaining why SOM
694 and litter are lower even with lower bacteria and fungi. These results also indicate that the model
695 runs reach an equilibrium in C pools, soil porosity and hydrology after ca. 1000 days.

696

## 697 Conclusions

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

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

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

721 In conclusion, KEYLINK is a first step towards a new generation of ecosystems models that
722 include functional diversity, trophic structures and ecological processes as important factors

723  shaping soil/ecosystem carbon and water cycling. Future versions, fed by more detailed data, will
724  need to be developed in order to improve our current predictive capacity.

725

## Acknowledgements

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

730

## References

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

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

740  Campbell. E. E., Parton, W. J., Soong, J. L., Paustian, K., Hobb, N. T., and Cotrufo, M. F.
741      (2016). Using litter chemistry controls on microbial processes to partition litter carbon
742      fluxes with the Litter Decomposition and Leaching (LIDEL) model. *Soil Biology and*
743      *Biochemistry, 100*, 160-174.

744  Campbell, E. E., and Paustian, K. (2015). Current developments in soil organic matter modeling
745      and the expansion of model applications: a review. *Environmental Research Letters,*
746      *10*(12), 123004.

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

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

752  Chuine, I. (2000). A unified model for budburst of trees. *Journal of Theoretical Biology, 207*(3),
753      337-347.

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

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

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

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

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

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

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

775 Geisen, S., Briones, M. J., Gan, H., Behan-Pelletier, V. M., Friman, V.P., de Groot, G. A.,
776         Hannula, S. E., Lindo, Z., Philippot, L., Tiunov, A. V., and Wall, D. H. (2019). A
777         methodological framework to embrace soil biodiversity. *Soil Biology and Biochemistry,*
778         *136*, 107536.

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

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

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

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

790 Horemans, J., Roland, M., Janssens, I., and Ceulemans, R. (2017). Explaining the inter-annual
791         variability in the ecosystem fluxes of the Brasschaat Scots pine forest: 20 years of eddy
792         flux and pollution monitoring. In: *EGU General Assembly Conference Abstracts, 19*,
793         10402.

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

797 Janssens, I. A., Sampson, D. A., Curiel Yuste, J., Carrara, A., and Ceulemans, R. (2002). The
798         carbon cost of fine root turnover in a Scots pine forest. *Forest Ecology and Management,*
799         *168*(1-3), 231-240.

800    Jones, C. G., Lawton, J. H., and Shachack, M. (1994). Organisms as ecosystem engineers. *Oikos*
801        *69*, 373-386.

802    Jörgensen, S. E. (2009). *Ecological modelling: an introduction*. WIT press. ISBN 978-1-84564-
803        408-6

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

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

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

809    Lavelle, P., Spain, A., Blouin, M., Brown, G., Decaëns, T., Grimaldi, M., Jiménez, J. J., McKey,
810        D., Mathieu, J., Velasquez, E., and Zangerlé, A. (2016). Ecosystem engineers in a self-
811        organized soil: a review of concepts and future research questions. *Soil Science, 181*(3/4),
812        91-109.

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

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

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

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

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

826    Persson, T., Bååth, E., Clarholm, M., Lundkvist, H., Soderstroem, B. E., and Sohlenius, B.
827        (1980). Trophic structure, biomass dynamics and carbon metabolism of soil organisms in
828        a Scots pine forest. *Ecological Bulletins, 32*, 419-459.

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

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

835    Saxton, K. E., Rawls, W., Romberger, J. S., and Papendick, R. I. (1986). Estimating generalized
836        soil-water characteristics from texture 1. *Soil Science Society of America Journal, 50*(4),
837        1031-1036.

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

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

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

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

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

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

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

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

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

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) |

Calibration data.

Data of C pools (see **Table 1**) used for the model calibration. 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$^{-3}$) |
|---|---|---|
| $B_b$ | 15.1 | 3.02 |
| $B_f$ | 15.1 | 3.02 |
| $B_{myc}$ | 160 | 32 |
| $B_{bvores}$ | 0.1 | 0.02 |
| $B_{fvores}$ | 0.8 | 0.16 |
| $B_{det}$ | 0.6 | 0.12 |
| $B_{eng}$ | 0.2 | 0.04 |
| $B_{hvores}$ | 0.2 | 0.04 |
| $B_{pred}$ | 0.4 | 0.04 |
| $L_{surf}$ | 2680 | 335 |
| SOM | 11470 | 1433.75 |

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 |

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** | $1.970 \pm 0.424$ |
| **Fungi** | $0.295 \pm 0.134$ |
| **Mycorrhiza** | $2.208 \pm 0.302$ |
| **Bacterivores** | $0.205 \pm 0.098$ |
| **Fungivores** | $0.095 \pm 0.050$ |
| **Detritivores** | $0.091 \pm 0.050$ |
| **Engineers** | $0.292 \pm 0.062$ |
| **Herbivores** | $0.028 \pm 0.018$ |
| **Predators** | $0.408 \pm 0.060$ |

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

Averages and standard deviations 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% | C:N$_{lit}$ 40 | pH 5.9 | B$_{pred}$ 0 | clay 15% |
|---|---|---|---|---|---|---|
| **Bacteria** | 5.98±36.66 | 6.94±41.01 | 6.35±37.85 | 5.90±36.42 | 2.67±20.51 | 4.75±35.11 |
| **Fungi** | 210.14±339.79 | 224.83±337.38 | 231.33±347.78 | 158.06±320.83 | 30.70±197.08 | 141.83±278.65 |
| **Mycorrhiza** | 39.83±20.62 | 40.20±21.63 | 40.03±20.71 | 38.70±20.59 | 29.75±19.26 | 39.27±18.45 |
| **Bactvores** | 4.63E-04±5.15E-03 | 4.65E-04±5.16E-03 | 4.63E-04±5.15E-03 | 2.24E-04±1.44E-02 | 2.88E-02±6.13E-01 | 5.04E-04±5.33E-03 |
| **Fungivores** | 0.30±3.73 | 0.31±4.05 | 0.39±4.59 | 0.18±2.86 | 0.90±10.29 | 0.40±5.67 |
| **Detritivores** | 2.49±36.04 | 2.12±30.25 | 2.01±30.62 | 3.83±96.81 | 145.39±363.51 | 0.87±10.06 |
| **Engineers** | 0.04±0.72 | 0.04±0.57 | 0.04±0.70 | 0.51±7.49 | 1.54±5.69 | 0.05±0.60 |
| **Herbivores** | 0.12±1.80 | 0.12±1.72 | 0.13±1.81 | 0.02±0.54 | 5.48±9.39 | 0.16±2.09 |
| **Predators** | 3.33±38.90 | 2.86±30.86 | 2.87±33.47 | 6.41±93.12 | 0.00±0.03 | 1.36±10.80 |
| **Litter** | 3695.08±1732.03 | 3262.71±1618.91 | 3481.79±1713.76 | 3728.85±1622.88 | 2727.81±1854.83 | 4245.54±1876.07 |
| **SOM** | 10825.04±3210.90 | 10941.01±3236.18 | 10742.97±3213.22 | 9347.23±3586.95 | 3175.08±2224.52 | 11655.70±2676.10 |

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

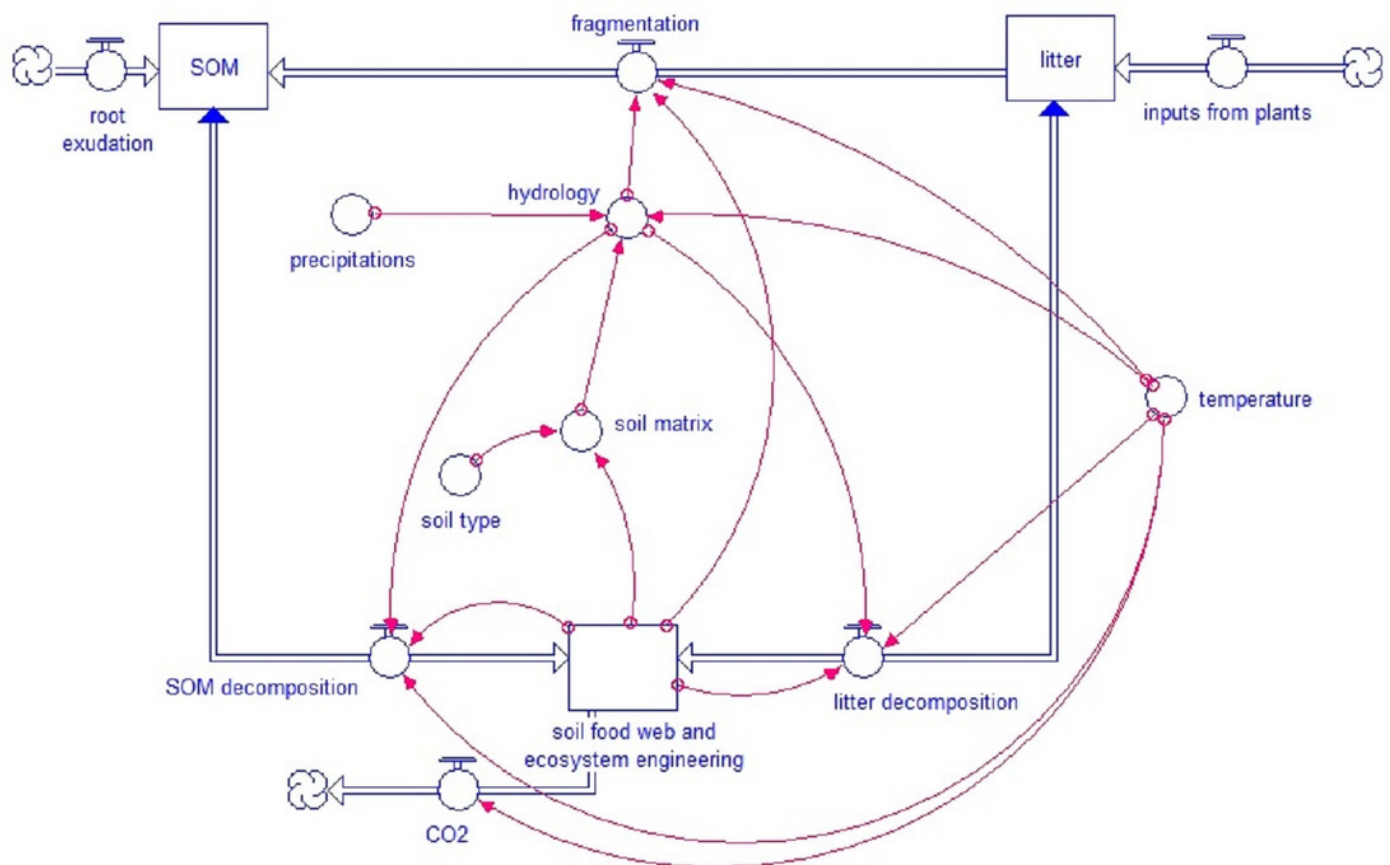
posterior distribution, with a basal scenario using reference input parameters

(**Supplemental File S2**), and the same changes from it as in **Table 5**.

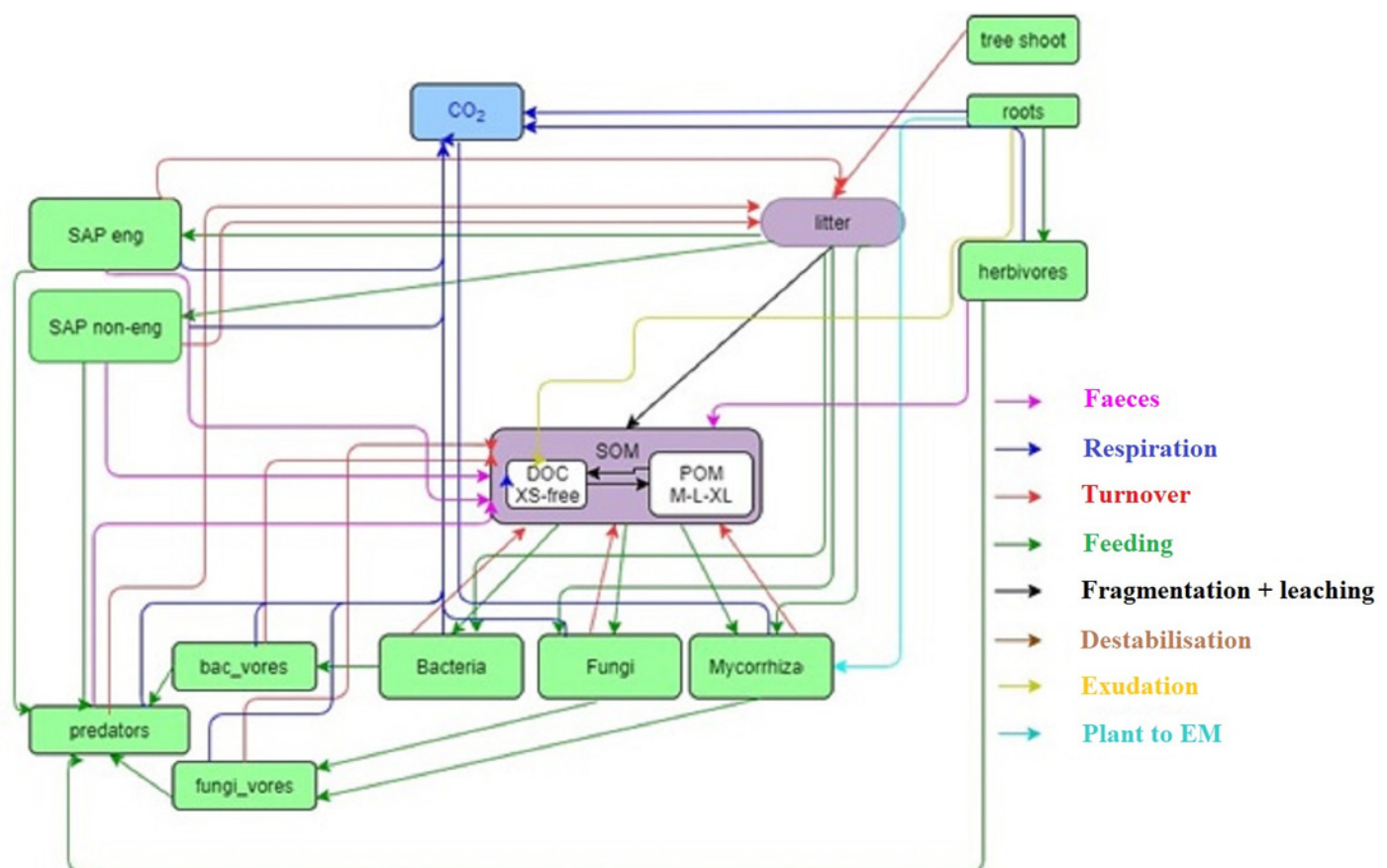
| | units | basal | rec 20% | C:N$_{lit}$ 40 | pH 5.9 | B$_{pred}$ 0 | clay 15% |
|---|---|---|---|---|---|---|---|
| **R$_{bact}$** | g C m$^{-3}$ | 250.2±222.8 | 293.6±227.2 | 262.3±225.7 | 248.9±222.4 | 119.3±149.9 | 204.5±225.9 |
| **R$_{fun}$** | g C m$^{-3}$ | 3924.8±5043.6 | 4190.0±5035.6 | 4215.4±5166.1 | 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 |

# Figure 1

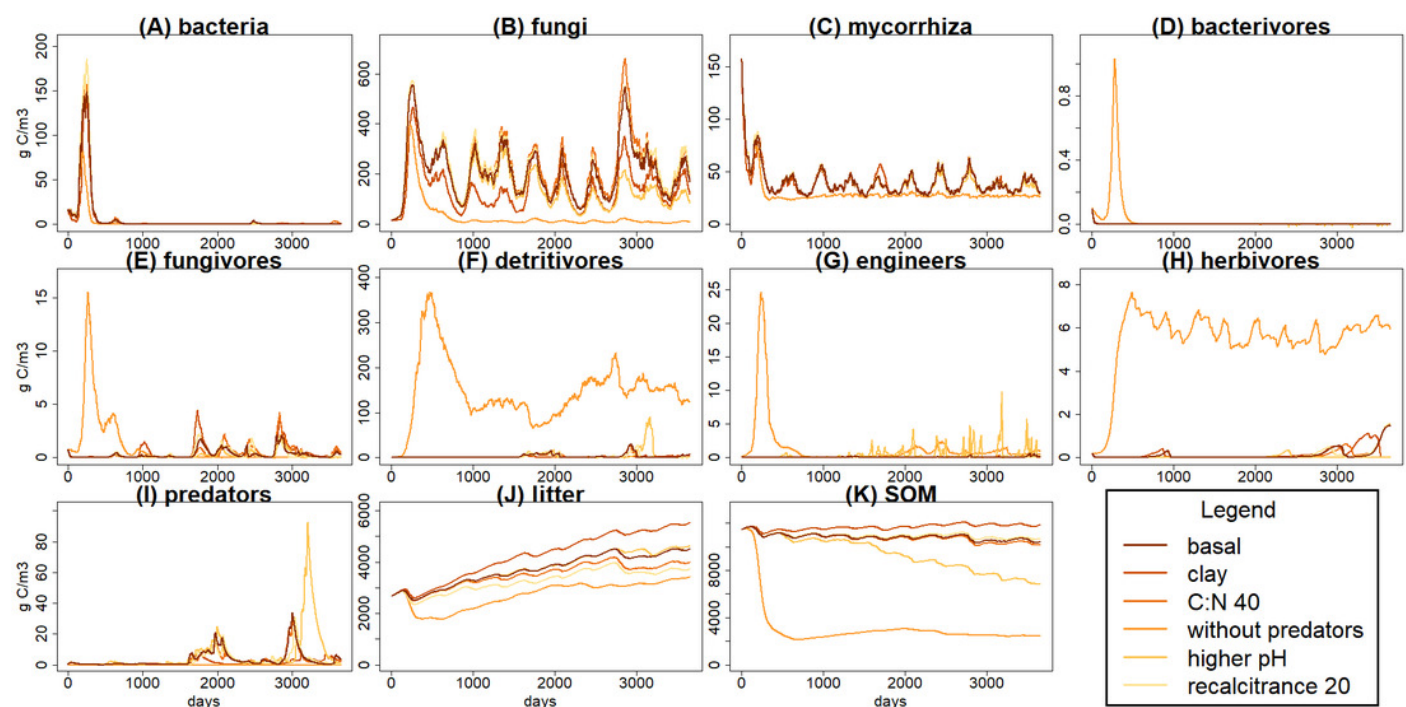
Simplified model scheme.

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 (food web, litter and SOM) 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. POM is particulate organic matter (medium (M), large (L) and extra large (XL) sizes), and DOC is dissolved organic carbon (extra small (XS) or free). Engineer species are represented as "SAP eng", while "SAP non-eng" are the detritivores. EM stands for ectomycorrhiza.

# Figure 3

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), and the alternative scenarios: higher clay content (red), lower input litter C:N (yellow), excluding predators (dark blue), higher soil pH (green) and lower litter recalcitrance (light blue). 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

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 3**). 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.