Dynamic characterisation of SOM degradation with an aerobic reaction network

A mechanistic toolbox

by

Marilou Sleiderink

to obtain the degree of Master of Science
in Civil Engineering
at the Delft University of Technology,
to be defended publicly on Friday September 16, 2016 at 1:00 PM.

Student number:	4409329	
Project duration:	September 24, 2015 – September 16, 2016	
Thesis committee:	Prof. dr. ir. T.J. Heimovaara,	TU Delft, head of committee
	Prof. dr. ir. M. de Kreuk,	TU Delft, external supervisor
	Dr. S.J. Laumann,	TU Delft, supervisor
	Ir. A.G. van Turnhout,	TU Delft, supervisor
	Ir. J. Zhou.,	TU Delft, supervisor

TU Delft Delft University of Technology

Abstract

The growing agreement that soil organic matter (SOM) is an essential factor to predict and optimise a wide range of soil ecosystem services, has made the lack of generally accepted mechanistic modelling tools for SOM degradation a pressing matter. In this study, we developed a state of the art mechanistic toolbox which embodies the core of SOM degradation, while maintaining a model structure that allows flexible addition of interactions with other soil components. The theoretical framework of the toolbox is an aerobic reaction network where SOM is defined as a mixture of known organic compounds from plant and microbial origin, that interact with microorganisms in a non-limiting aqueous environment. Simulations of the toolbox show that the bulk properties of SOM can be interpreted as weighted averages of the properties of individual organic compounds. The downside of implementing a novel theoretical approach in the toolbox, is that some crucial parameter values are poorly documented in literature. To tackle this issue, we incorporated a Bayesian inference tool, which is capable of selecting an experimental design that makes optimal use of the inherent model structure to maximise the future parameter identifiability. The toolbox shows promising outlooks to both a) increase the acceptance of the state of the art insights in SOM degradation by reinterpreting older experimental findings, while b) further enhancing mechanistic modelling of interaction processes between SOM and other soil components to, ultimately, make accurate predictions of SOM degradation within a complex soil environment.

Contents

Preface

As a university student, I have pursued a bachelor degree in bioscience engineering at the University of Leuven and then a master degree in geo-engineering of the Delft University of Technology. When switching majors, I did not suspect to see those five years of studies culminating into a master thesis which so fittingly combines both worlds through the field of soil biogeochemistry. The most rewarding aspect of my thesis was that I started my research at a fundamental level and then still saw it come together in a concrete, and almost tangible, numerical model, in a bit less than a year. At times it felt like more than I bargained for, but the final result is definitely worth it. Here is to hoping that this outcome will contribute in its own way to the field of soil organic matter.

I would like to thank Prof. dr. ir. Timo Heimovaara for sharing his passion for this research topic, as well as giving me the freedom to explore the subject through my personal approach. I have greatly appreciated that Prof. Dr. Ir. Merle de Kreuk has provided much needed feedback from a fresh pair of eyes. I would also like to thank Dr. Susanne Laumann, Ir. André van Turnhout and Ir. Jiani Zhou for being so invested in my thesis and its outcome. Our often intense meetings have played a key role in delivering a thesis which is as straightforward as it could be for a topic that has been called "the most vague field in soil science" (Guggenberg et al., 2005).

While a thesis is largely an exercise in independently conducting research, I especially appreciated everyone who found their own way to show me that I was *not* alone in this challenge. Among those people are my boyfriend Tom, my parents Remco and Kathleen, my sister Nora and close friends both in Delft and in Leuven. The times you went to the nightshop to get me comfort food? When you let me crash on your couch on my numerous visits in Delft? Enduring my frustrations over a cup of coffee? Pizza at the geo-corner? Reading and rereading my thesis? Those are the things that matter and which I will fondly look back upon when remembering the last year of my life as a student. Thank you!

Marilou
Between Delft and Leuven
13 September 2016

Abbreviation list

$\%_{ArO}$	Percentage β-aryl ether bonds in lignin
$\%_{BiPh}$	Percentage biphenyl bonds in lignin
$\%_{C_{16}}$	Percentage C_{16} LCFA in cutin
$\%_{C_{18}}$	Percentage C_{18} LCFA in cutin
$\%_{C_6}$	Percentage hexose in hemicellulose
$\%_{chit}$	Percentage chitin in fungal necromass
$\%_{fungi}$	Percentage of fungal biomass
$\%_{G}$	Percentage Guaiacyl units in lignin
$\%_{G}$	Percentage P-Hydroxyphenol units in Lignin
$\%_{glucan}$	Percentage glucan in fungal necromass
$\%_{pept}$	Percentage peptidoglycan in bacterial necromass
$\%_{prot,b}$	Percentage protein in bacterial necromass
$\%_{prot,f}$	Percentage protein in fungal necromass
$\%_{S}$	Percentage Syringyl units in lignin
$\mathbf{\Delta G_f^\circ}$	Gibbs free energy of formation
η_{Ac}	Acetylation degree of hemicellulose
ASM	Activated Sludge Model
ATP	Adenosine Triphosphate
BIC	Bayesian Information Criterion
$\mathbf{C_{16}}$	LCFA with a NoC of 16
$\mathbf{C_{18}}$	LCFA with a NoC of 18
C/N	Carbon-to-Nitrogen Ratio
$\mathbf{D_{kl}}$	Kullback-Leibler Divergence
DIR	Direct Hypothesis
EPS	Extrapolymeric Substances
G	Guaiacyl unit
G	P-Hydroxyphenol unit
LCFA	Long Chain Fatty Acids
MGE	Microbial Growth Efficiency
MIT	Mineralisation-Immobilisation Turnover
MOM	Microbial Organic Matter
NAG	N-Acetylmuramic Acid
NAM	N-Acetylglucosamine
NMR	Nuclear Magnetic Resonance
NoC	Number of Carbon Atoms
PAR	Parallel Scheme
RN	Reaction Network
s	Syringyl unit
SD	Standard Deviation
SOC	Soil Organic Carbon
SOM	Soil Organic Matter
VFA	Volatile Fatty Acids
w/v	Weight over Volume Percentage Concentration

List of Figures

List of Tables

1 Introduction

Soils provide a wide range of ecosystem services, as the result of complex interactions between several soil components. Soil organic matter (SOM) has been known to be a crucial factor for soil fertility for decades [1, 2]. Nowadays, SOM is identified as playing an essential role in an increasing number of ecosystem services, such as water quality, erosion resistance and climate change mitigation [1]. Not surprisingly, extensive time and effort have been dedicated towards predicting and, ultimately, optimising the ecosystem services for their end-goal.

By consequence, there is an urgent need for tools that can answer the ever-growing amount and variation in research questions regarding SOM [3]. Currently, the available numerical models are falling short of this demand due to their limited applicability. The prevailing models, such as RothC [4], are constrained to the turnover of carbon and sometimes nitrogen in the soil, without the possibility to extract information on additional properties of SOM [5]. These models are developed based on data collected from field studies [4]. Due to this empirical nature, the validity of their predictions is restricted to the range of observations used upon building the model [1, 4]. For instance, if soil temperatures increase drastically due to climate change, the current models cannot be used to predict how carbon turnover will react within this new temperature range [1]. Overall, using empirical modelling to answer the vast diversity in SOM-related research would require an individual model for each ecosystem service, which would then still encounter major difficulties when making predictions in diverging conditions. Mechanistic modelling, on the other hand, aims to make predictions by mimicking the processes leading up to the outcome [6]. Given that the ecosystem services are all merely different outcomes of the same interaction processes, mechanistic modelling is a much more efficient approach to SOM predictions and optimisation over a range of scientific fields. Unfortunately, generally accepted mechanistic numerical models of SOM degradation are lacking to date.

A major prerequisite of mechanistic modelling is knowledge of the processes in question, which is where the problem lies with modelling of SOM degradation. The fact that mechanistic models failed to develop, can be traced back to a longstanding misconception of the process of SOM degradation. Traditionally, it has been thought that the majority of SOM consists of humic substances [7]. Humic substances constitute of humic acids, fulvic acids and humin [8]. The constituents can be divided through a sequential extraction scheme of acidic and alkaline solutions, which operationally defines each fraction [8]. Humic and fulvic acids were thought to be products of complex condensation reactions between disassembled plant material, resulting in *de novo* macromolecules. The macromolecules were expected to be refractory with regard to degradation by microorganisms, allowing them to persist in the soil for periods several times longer than the original plant material [9]. However, the macromolecules in question had never been directly observed in samples of SOM [10]. Their absence was disregarded as merely an analytical artefact, by assuming the analysis methods were interfering with the actual structure of the humic substances [10].

In the last ten years, a new interpretation for SOM degradation has been evolving from two key insights, namely that a) neoformation of macromolecules is negligible, and b) that SOM persists in the soil not solely due to its own refractory nature, but due to the combined action of several stabilisation mechanisms [1, 2, 7, 11]. The first insight is the direct result of the application of advanced nuclear magnetic resonance (NMR) spectroscopy to identify traditional fractions of humic substances [10]. The measured signals show that the vast majority of SOM mixtures can be explained by organic molecules of plant and microbial origin [10]. This observation does not

exclude the existence of *de novo* macromolecules, but strongly refutes their importance [10]. The second insight opposes the notion that certain microbial and plant compounds persist for long periods of times, by virtue of their intrinsic refractory nature [1,7]. Laboratory experiments show that the degradation rate of even the most resistant plant compounds is still orders of magnitude higher than the average turnover rate of organic matter in the soil. Rather, the degradation rate of SOM is governed by environmental factors, which protect or stabilise SOM in various ways against their degradation by microorganisms [1, 7]. The main mechanisms are aggregate formation, mineral binding and metal binding [1, 7].

To the author's knowledge, only two attempts at translating the new theoretical insights into a mechanistic numerical model have been published. The first attempt is the AggModel by Segoli et al. (2013) [3]. The AggModel focuses on the second insight, by modelling aggregate dynamics, SOM degradation and mineral binding of SOM as mechanistic processes. The organic matter itself is characterised according to its state of association with minerals and aggregates. The second attempt is the BAMS1 model by Riley et al. (2014) [12]. BAMS1 includes the first insight by developing a reaction network for SOM degradation by defining SOM as a mixture of carbon compounds of plant and microbial origin. The second insight is integrated in the model by assigning relatively fast reaction rates to the degradation of these compounds. Furthermore, SOM is then protected by its interaction with soil minerals.

First and foremost, the goal of this paper is to create a numerical model that embodies solely the core of SOM degradation, but does feature the possibility to add interactions with other soil components in a mechanistic manner in later applications. Concretely, this goal is pursued by developing a reaction network based on state of the art insights into SOM. An extensive literature research is conducted to define SOM entirely as known compounds of plant and microbial origin. The degradation of SOM is the direct result of interaction of SOM with microorganisms and soil water. By consequence, microorganisms and soil water are inherent components of the SOM degradation core and included in the model. Furthermore, the need to use the core model to add interactions with other soil components is met by the characterisation of the SOM compounds with various properties. The reaction network is implemented in a numerical forward model.

The main drawback of the forward model developed in this paper, is that it relies on parameter values which are poorly documented in current literature. For this reason, a second research objective is to create a tool that can suggest the design of an experiment which would make optimal use of the inherent model sensitivity to identify the value of the most uncertain parameters. To conduct this analysis, the forward model is expanded with the DREAM algorithm of Vrugt et al. (2016) as implemented by Van Turnhout et al. (2016). The third and last research objective is to allow customised adjustments of the presented reaction network, using objective assessment criteria. To achieve this, we use the DREAM module for network selection, as done in Van Turnhout et al. (2016).

The findings of the literature research are presented in Section 2. The development of a theoretical reaction network based on the literature findings is explained in Section 3. Section 4 addresses the implementation of the reaction network into a toolbox. Section 5 describes a single run of the forward model with baseline values and an analysis to assess the suitability of experimental designs for uncertain parameters quantification (Section 5.3). The results of both methods are presented in Section 6, and subsequently discussed in 7. Finally, the conclusions of this paper are presented in Section 8.

2 Literature Research

This section presents the findings from the literature study which are relevant for the further development of a reaction network for SOM degradation. First, we outline the main concepts of SOM degradation, as illustrated in Figure 1. Afterwards, the SOM compounds of plant origin involved in aerobic degradation are described. Subsequently, the SOM of microbial origin is defined as well. Finally, the kinetics of these reactions are determined. The section is concluded with a brief description of anaerobic SOM degradation. This description highlights the significant difference with aerobic degradation, although developing a reaction network for anaerobic SOM degradation is beyond the scope of this paper.

2.1 Current view on SOM degradation

The reaction network is conceptually defined in a chronological order, starting with plant organic matter which enters the soil. The polymeric fraction of these plant residues is depolymerised by extracellular enzymes. The enzymes are produced by both major groups of living biomass in the soil, i.e fungi and bacteria [13]. The depolymerisation yields several monomers. Monomers function as substrates for microbial growth, upon which carbon dioxide (CO_2) and other byproducts such as extrapolymeric substances (EPS) are produced [13]. The efficiency with which substrates are build up as biomass and byproducts is called the microbial growth efficiency (MGE) [13]. Both depolymerisation and monomer uptake occur at a rate linked to the intrinsic resistance of the compounds. Stabilisation mechanisms act as barriers between SOM and the (extra)cellular enzymes, which slow down the overall reaction [1, 7]. The total amount of living biomass is the result of both growth and decay. Decayed biomass is called necromass, which consists of microbial residues, mostly polymers [13]. These polymers undergo the same process as polymers from plant origin. Without input of fresh plant material, the microbial SOM fraction will grow at the expense of plant SOM. This concept is summarised in Figure 1. The production of microbial byproducts is not dealt with in this research. The barriers or stabilisation mechanisms are not part of this reaction network, which solely focuses on a dynamic way to represent SOM as a component (or module) within a broader soil system (or model). Section 2.2 will identify the major plant compounds and their reaction paths. Section 2.3 then describes the decay of living biomass and the subsequent degradation of microbial organic matter (MOM). The kinetics of all the above reactions are discussed in Section 2.4. Section 2.5 highlights the limitations of this literature research.

2.2 Aerobic degradation of plant residues

Defining the SOM fraction of plant origin is structured in two steps. The first step, described in Section 2.2.1, characterises all individual polymer types through one representative compound, one chemical formula and one or more monomers. In Section 2.2.2, the depolymerisation of these polymers is addressed. Depolymerisation by hydrolysis releases the same monomers as for polymerisation, as opposed to oxidative depolymerisation which can yield different monomers. Section 2.2.3 describes the monomeric uptake, which leads to the removal of the plant residues from the SOM mixture.

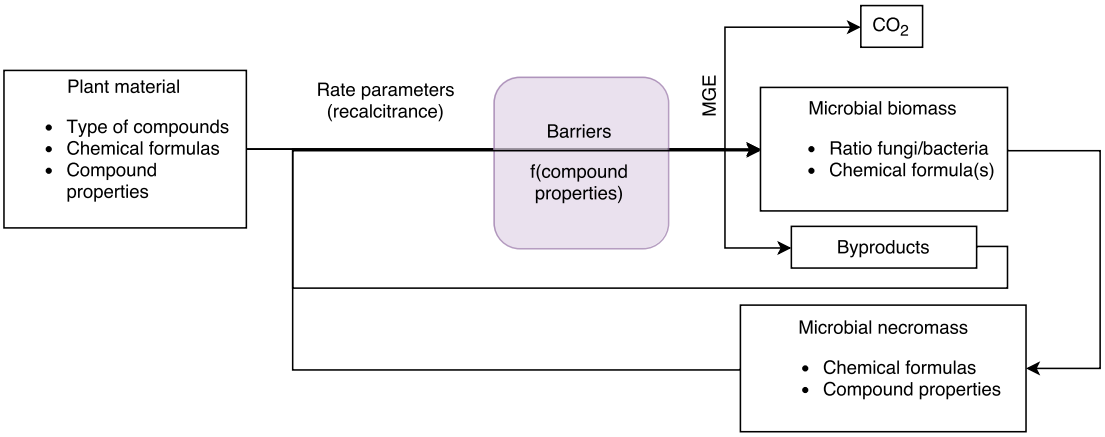


Figure 1: Conceptual model of SOM degradation

2.2.1 Input of plant residues

In this section we determine which plant residue polymers are present in the soil. These residues enter the soil system both as above ground input and as below ground input through root exudates [7]. Plants consist of eight major types of polymers, i.e. starch, cellulose, hemicellulose, lignin, tannins, cutin, suberin and proteins [14]. Additionally, lipids are prominently present in plant organic matter, although they are not polymers. The eight polymer types are first defined in function of their role within the plant. Afterwards, the polymers are further described in function of their monomers and linkage among those monomers. The monomers and linkage types are always unequivocally defined, with a fixed chemical formula and chemical properties. For certain polymers extra parameters are introduced to characterise the variability with regard to monomers and linkage. With the help of the previously defined monomers, linkage and possible extra parameters, the polymers can be characterised through a clear expression of their chemical formula, and their functional groups. The functional groups discussed in this literature research are carboxyl groups, other carbonyl groups and phenolic groups. The functional groups of the polymer can again be derived from its monomers and their linkage. All monomers defined in this section are included in Table 1. All polymers defined in this section are included in Table 2 with an expression for their chemical formula. In the chemical formulas of the polymers, the n subscript represents an arbitrary polymerisation degree. Table 3 contains the characterisation of the plant polymers based on their functional groups.

1. Starch is a storage polysaccharide in vascular plants [14]. It is composed of amylose and amylopectin, two different polysaccharides of hexose monomers ($C_6H_{12}O_6$). The hexose in question is glucose. The glucose monomers are linked together by glycosidic bonds. The glycosidic bond is created by a dehydration reaction, which releases one H_2O molecule when linking two monomers. Therefore, the chemical formula of starch is $(C_6H_{10}O_5)_n$. Glucose does not have any carboxyl, carbonyl or phenolic groups. The glycosidic bond does not add any of these either, so starch does not have any carboxyl, carbonyl or phenolic groups.

2. Cellulose is the principal structural component of plant cells [14]. It is a polysaccharide which consists of glucose monomers [8]. The glucose units are linked through $\beta-(1-4)-$glycosidic bonds. The glycosidic bond is created by a dehydration reaction, which

releases one H_2O molecule when linking two monomers. The chemical formula of cellulose is therefore $(C_6H_{10}O_5)_n$. Glucose does not have any carboxyl, carbonyl or phenolic groups. The glycosidic bond does not add any of these either, so cellulose does not have any carboxyl, carbonyl or phenolic groups.

3. Hemicellulose is the second most common structural polysaccharide [14]. Hemicellulose has a much higher degree of heterogeneity than cellulose. Hemicellulose is a complex branched structure made up of a variety of monosaccharides [8], mainly pentoses ($C_5H_{10}O_5$) and some hexoses ($C_6H_{12}O_6$) [15]. The percentage of hexose is indicated with $\%_{C_6}$. The percentage of pentose is thus $1-\%_{C_6}$. Hemicellulose has several substitutes on its side chains, mainly acetic acid ($C_2H_4O_2$) [15]. The percentage of monosaccharide monomers which are acetylated can be indicated by the acetylation degree η_{Ac}. The acetylation of a hydroxyl group on the monosaccharide removes one H_2O molecule from the monosaccharide and the acetic acid in total. A general expression for the chemical formula of hemicellulose is thus $(C_{5+\%_{C_6}}H_{10+2\%_{C_6}-2(1+\eta_{Ac})}O_{5+\%_{C_6}-1(1+\eta_{Ac})})_n$. For example, xylan is one of the most common hemicelluloses [14]. Xylans have different acetylation degrees [16]. Xylans originating from dicots are expected to have one acetic acid group per two xylose units [16]. Those hemicellulose polymers thus have a chemical formula $(C_6H_9O_{4.5})_n$. Hexose and pentose monosaccharide do not have any carboxyl, carbonyl or phenolic group. As mentioned for cellulose, the glycosidic bond does not add any functional groups either. However, for each acetylated monosaccharide hemicellulose gains one carboxyl group. Therefore, hemicellulose has η_{Ac} carboxyl groups.

4. Lignin provides structural rigidity to a plant by filling out cell walls [14]. It is a complex, highly aromatic structure [8, 14]. The monomers are phenylpropanoid units, i.e. guaiacyl (G, $C_{10}H_{12}O_3$), syringyl (S, $C_{11}H_{14}O_4$) and p-hydroxyphenol (P, $C_9H_{10}O_2$) units [8, 17]. The chemical formula of the lignin polymer depends on the ratio of G:S:H and on the linkage between the monomers [17]. The parameter $\%_G$ indicates the percentage of phenylpropanoid units that are G units. The percentage of S units is indicated with $\%_S$. The percentage of P units can be calculated from the previous ones, i.e $1-\%_G-\%_S$. In this literature research we examine the β-aryl ether bond, the biphenyl linkage and the phenylcoumarane linkage, which are the most common bond types in lignin [17]. This examination is conducted for the polymeratisation of a p-hydroxyphenyl unit (Figure 2). For the β-aryl ether bond, two times a hydrogen radical is removed and then a H_2O molecule is added [17]. The result is therefore that the polymer contains one oxygen atom more than its constituting monomers. For the biphenyl bond two times a hydrogen radical is removed [17]. The polymer then contains two hydrogen atoms less than the monomer. For the phenylcoumarane linking, two hydrogen radicals are also removed when linking two monomers [17]. The biphenyl bond and the phenylcouramane linking therefore have the same effect on the chemical formula of the polymer. We assume that these binding mechanisms are the same for linking other monomeric units as well. The percentage of bonds which are β-aryl ether bonds is indicated with $\%_{ArO}$. The other bonds are assumed to be biphenyl or phenylcoumarane bonds. A general expression for the chemical formula of lignin can thus be set up with parameters $\%_G$, $\%_S$ and $\%_{ArO}$. The general expression is given by $(C_{9+1\%_G+2\%_S}H_{10+2\%_G+4\%_S-2(1-\%_{ArO})}O_{2+1\%_G+2\%_S-\%_{ArO}})_n$. Characterizing lignin by its functional groups is again done through its units and the linkages. All units have one phenolic functional group. Both the phenylcoumarane linkage and the aryl-ether linkage remove the phenolic functional group (transforming it respectively into an acyclic ether and a cyclic ether). The biphenyl linkage leaves the phenolic functional group unchanged. To characterise the functional groups of lignin, we therefore need one additional parameter.

The percentage of biphenyl linkage is indicated by $\%_{BiPh}$. The amount of phenolic groups in lignin then becomes $\%_{BiPh}$.

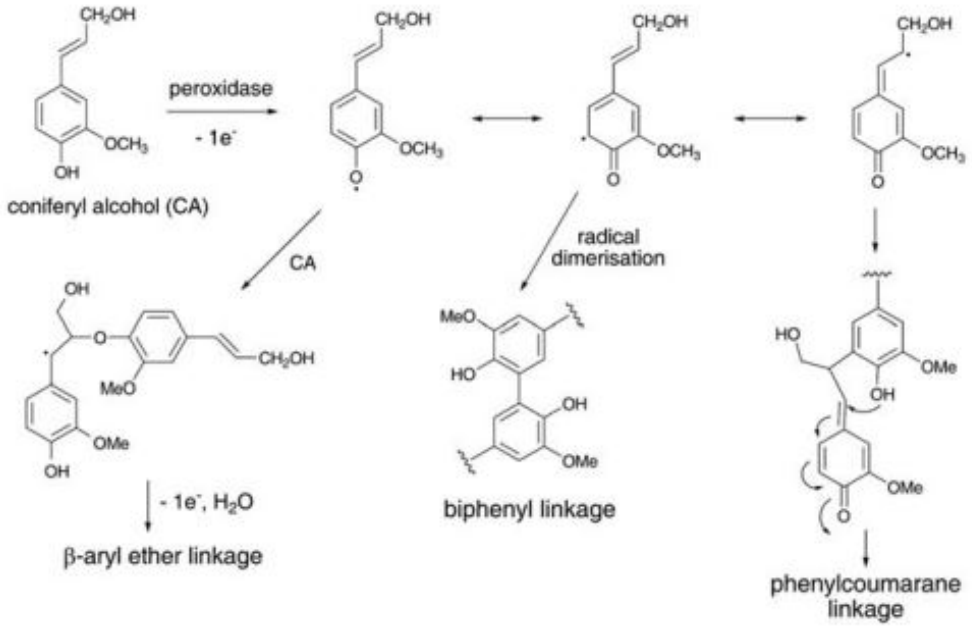


Figure 2: Formation of bonds in lignin (*Source: Bugg et al (2011)*)

5. Tannins are polyphenols that can be found in higher plants [14]. Tannins are capable of precipitating proteins from solutions, through non-reversible reactions [18]. They can be divided into condensed tannins and hydrolysable tannins. Hydrolysable tannins consist of two units, i.e. sugars (glucose or similar) and phenolic acids [14]. The monomers of condensed tannins are flavan-3-ols, which have two aromatic rings [14]. The most common flavan-3-ol for tannin is catechin ($C_{15}H_{14}O_6$) [14]. The monomers are bound through carbon-carbon bonds [14]. The polymer can therefore be represented as $(C_{15}H_{12}O_6)_n$. Catechin has four phenolic groups. The carbon-carbon bond does not change these functional groups. Therefore, tannin also contains four phenolic groups.

6. Cutin forms the macromolecular frame of the plant cuticle [14]. Together with the waxes and lipids in the cuticle, it is the principal source of aliphatic materials in soil [19]. It is an insoluble polyester of cross-linked hydroxy-fatty acids and hydroxyepoxy-fatty acids, which are long chain fatty acids (LCFA) [19]. Cutin monomers can be divided based on their chain length, which is either C_{16} or C_{18} [14]. The major C_{16} component is 10,16-dihydroxypalmitic acid ($C_{16}H_{30}O_4$) [20]. The major C_{18} component is 18-hydroxyoleic acid ($C_{18}H_{30}O_4$) [20]. The proportion of C_{16} to C_{18} monomers can vary greatly. For instance, cutin of fast-growing plants contains mostly C_{16} monomers, whereas the cutin of an apple cuticle contains both types of monomers [20]. We introduce the parameter $\%_{C_{18}}$ to indicate the percentage of monomers with chain length C_{18}. The monomers are bound through ester linkages upon dehydration of the monomers [20]. Further research on the amount of linkage is necessary, but it seems justified to assume that each monomer is bound to two other monomers by two ester linkages. A general expression for the chemical formula

of cutin can be given by $(C_{16+2\%_{C_{18}}}H_{28}O_3)_n$. The monomers 10,16-dihydroxypalmitic acid and 18-hydroxyoleic acid both contain one carboxyl group. Through the ester linkage, the carboxyl group is transformed into a carbonyl group in cutin.

7. Suberin is a cell wall component of cork cells and its content is high in bark and plant roots [14]. It is a polyester, like cutin, but in addition to aliphatic monomers, it contains aromatic components [14, 20]. The aliphatic monomers have chain lengths of C_{20} to C_{30}, which is longer than cutin monomers [14]. The aromatic components are phenolic acids [14]. Suberin is also a polyester present in the cuticle of vascular plants [14]. The aliphatic monomers and aromatic monomers form distinct domains [14]. The composition and structure of suberin is insufficiently known to set up a general expression for its chemical formula.

8. Proteins have varying roles in the plant, ranging from enzymes to storage substances [14]. They are polypeptides, i.e long chains of amino acids, made up of approximately 20 common amino acids [14]. Amino acids differ from one another in chemical formula and chemical properties. We therefore present a different approach for setting the chemical formula of proteins and amino acids. Torabizadeh et al (2011) [21] states that, even though proteins have a large range of functions and structures, it is possible to find a general chemical formula. The suggested formula is $(CH_{1.58}N_{0.28}O_{0.3}S_{0.01})_x$, where x represents any number of carbon atoms. To set a general expression for the chemical formula of protein which is conform to the previous ones, the number of carbon atoms should be that of the number of carbon atoms in the monomers. Amino acids typically have between 3 and 5 carbon atoms [21]. By setting the number of carbon atoms to four, the chemical formula of the protein becomes $(C_4H_{6.32}N_{1.12}O_{1.2}S_{0.04})_n$. A general chemical formula for amino acids is now backfigured from the polymerisation reaction of amino acids. Amino acids are linked together through peptide bonds. Peptide bonds are formed through a dehydration reaction. One H_2O molecule is thus removed when binding two amino acids. A general chemical formula for amino acids can thus be given by $C_4H_{8.32}N_{1.12}O_{2.2}S_{0.04}$. Amino acids consist of an amine functional group, a carboxyl group and a side chain. In this literature research we do not discuss amine groups. The side chains can have greatly varying functional groups. There is therefore always (at least) one carboxyl group in amino acids. The peptide bond removes the carboxyl group, which means only a carbonyl group remains in proteins, not regarding possible side chains.

9. Lipids are a heterogeneous group of plant compounds, which can fulfil different functions as well [14]. For instant, they are part of the plant cuticle which forms a water-repellent layer [14]. Plant lipids can be alkanes, alkenes, fatty acids and esters [14]. Due to their large variety, it is not possible to set a formula for their composition, nor for their functional groups. Nonetheless, it should be noted that lipids are not polymers, as opposed to the previously described plant compounds.

2.2.2 Depolymerisation of plant polymers

Once the plant residues enter the soil system, they are exposed to the action of extracellular enzymes from fungi and bacteria. These extracellular enzymes depolymerise the polymers into monomers. The depolymerisation occurs either through hydrolysis or oxidation of the monomeric bonds. Hydrolysis is the reverse reaction of dehydration. By consequence, the monomers released upon hydrolysis are identical to those described previously for polymerisation. On the

other hand, depolymerisation by oxidation is not necessarily the same as the reverse reaction of polymerisation. As a result, it is necessary to define new monomers in certain cases.

Of the eight polymer types described in Section 2.2.1, five can be hydrolyzed: starch, cellulose, hemicellulose, cutin and proteins. The hydrolysable bond and one example of an extracellular enzyme involved in the hydrolysis are discussed for each polymer. The monomers of starch and cellulose are linked by glycosidic bonds, which are hydrolysable by respectively amylase and cellulase [14, 22]. Hemicellulose monomers are also linked by glycosicid bonds, which can be hydrolyzed by, for example, xylanase [23]. The cutin monomers are bound through ester linkages, which can be hydrolyzed by cutinase [20]. The peptide bonds of protein can be hydrolyzed by protease [7, 19]. The hydrolysis reactants and products are summarised in Table 4, together with their stoichiometric coefficients.

The remaining three types of plant polymers mentioned in Section 2.2.1 require oxidative depolymerisation: lignin, condensed tannin and suberin. Suberin is however not further described due to lack of sufficient knowledge on its composition and depolymerisation mechanisms.

Lignin is mostly degraded by white-rot fungi [17, 23]. However, brown-rot fungi and even some bacteria can also degrade lignin through different pathways [17]. Several types of degradation products are found [17]. The most common breakdown products for fungal degraders are benzoic acids, with additional carboxy, methoxy and/or hydroxyl groups [17]. Typical benzoic acids are for instance dihydroxybenzoic acid ($C_7H_6O_4$), 4-hydroxy-3-methoxy-6-carboxybenzoic acid ($C_7H_8O_6$) and gallic acid, also known as 3,4,5-trihydroxybenzoic acid ($C_7H_6O_5$) [17]. Side-chains are cleaved by oxidation as well, which can form various volatile fatty acids (VFA) [17]. Laccases

Table 1: Monomer characterisation of plant residues

Monomer	**Chemical formula**	**-COOH**	**-Arom-OH**	**C=O**
Hexose	$C_6H_{12}O_6$	0	0	0
Pentose	$C_5H_{10}O_5$	0	0	0
Acetic acid	$C_2H_4O_2$	1	0	0
Guaiacyl unit	$C_{10}H_{12}O_3$	0	1	0
Syringyl unit	$C_{11}H_{14}O_4$	0	1	0
P-hydrophenyl unit	$C_9H_{10}O_2$	0	1	0
Catechin	$C_{15}H_{14}O_6$	0	4	0
10,16-dihydroxypalmitic acid (C_{16})	$C_{16}H_{30}O_4$	1	0	0
18-hydroxyoleic acid (C_18)	$C_{18}H_{30}O_4$	1	0	0
Amino acid	$C_4H_{8.32}N_{1.12}O_{2.2}S_{0.04}$	1	0	0

Table 2: Chemical formula of polymers in plant residues

Polymer	**Parameters**	**Chemical formula**
Starch	-	$(C_6H_{10}O_5)_n$
Cellulose	-	$(C_6H_{10}O_5)_n$
Hemicellulose	$\%_{C_6}$, η_{Ac}	$(C_{5+\%_{C_6}+\eta_{Ac}}H_{10+2\%_{C_6}-2(1+\eta_{Ac})}O_{5+\%_{C_6}-1(1+\eta_{Ac})})_n$
Lignin	$\%_G$, $\%_S$, $\%_{ArO}$	$(C_{9+1\%_G+2\%_S}H_{10+2\%_G+4\%_S-2(1-\%_{ArO})}O_{2+1\%_G+2\%_S-\%_{ArO}})_n$
Condensed tannins	-	$(C_{15}H_{12}O_6)_n$
Proteins	-	$(C_4H_{6.32}N_{1.12}O_{1.2}S_{0.04})_n$
Cutin	$\%_{C_{18}}$	$(C_{16+2\%_{C_{18}}}H_{28}O_3)_n$

are known to use the reduction of O_2 to H_2O for oxidation of lignin [17, 24]. The half reaction for O_2 is given by:

$$O_2 + 4H^+ + 4e^- \Rightarrow 2H_2O$$

Peroxidases such as lignin peroxidase (LiP) use the reduction of hydrogen peroxide (H_2O_2) to water [17, 24]. The half reaction is given by:

$$H_2O_2 + 2H^+ + 2e^- \Rightarrow H_2O$$

While the depolymerisation of hydrolysable tannins to its monomers, e.g. gallic acid and glucose, is explicitly mentioned in literature, this is not the case for the oxidative depolymerisation of condensed tannins. However, the uptake of flavan-3-ol as substrate for microorganisms is discussed with regard to condensed tannins [18]. It thus seems justified to assume that flavan-3-ol is a possible depolymerisation monomer.

The newly introduced monomers are summarised in Table 5. The stoichiometry of the oxidation reactions cannot be written with a general equation, as was the case for hydrolysis reactions, due to the larger variability in depolymerisation products. The half reaction for the polymer oxidative depolymerisation needs to be balanced with the half reaction of the reduction of the oxidator.

2.2.3 Substrate uptake of plant monomers

The depolymerisation of polymers by extracellular enzymes creates monomers which can function as substrates for microorganisms. The substrates are used in several metabolic reactions, which lead to biomass growth, maintenance and product formation [5]. A carbon source, a nitrogen source and water are crucial for these processes [5, 25]. Additionally, a reductor or an oxidator is necessary if the substrate is more, respectively less oxidated than the biomass itself [25]. The exact elemental composition of biomass differs depending on the type of fungi or bacteria. A commonly used elemental composition is $CH_{1.4}N_{0.2}O_{0.4}$ [6].

Add lipids The metabolic reactions can be split up in anabolic and catabolic reactions. The catabolic reactions break down the substrates into products, while storing energy in the form of adenosine triphosphate (ATP) [25]. In aerobic conditions, the main byproduct is CO_2 [25]. Anabolic reactions transform substrates into biomass components, using ATP [25]. Balancing the catabolic and anabolic reaction for biomass growth is achieved through the growth yield ($Y_{X/S}$). The growth yield of a specific substrate can be estimated from experimental data. Additionally,

Table 3: Functional groups of polymers in plant residues

Polymer	Parameters	-COOH	-Arom-OH	C=O
Starch	-	0	0	0
Cellulose	-	0	0	0
Hemicellulose	η_{Ac}	η_{Ac}	0	0
Lignin	$\%_{BiPh}$	0	$\%_{BiPh}$	0
Condensed tannins	-	0	4	0
Proteins	-	1	0	0
Cutin	-	0	0	1

Table 4: Depolymerisation of hydrolysable plant residue polymers

Polymer	**C-compounds**	**H_2O**
Starch	1 glucose	-1
Cellulose	1 glucose	-1
Hemicellulose	1 monosaccharide + η_{Ac} Acetic acid	-(1+η_{Ac})
Protein	1 Amino Acid	-1
Cutin	$\%_{C_{18}}C_{18}$ + (1-$\%_{C_{18}}$)C_{16}	-1

Kleerebezem et al. (2010) [25] have developed a method based on thermodynamic principles to determine the growth yield. However, the catabolic reaction is not always linked to an anabolic reaction, for instance when the catabolic reaction is used for maintenance purposes.

As mentioned before, a nitrogen source is also essential for microbial growth. Originally, it was thought that only nitrogen in its mineralised form, NH_3, could be used for biomass growth [5,26]. This hypothesis is called the Mineralisation-Immobilisation Turnover (MIT) scheme, where all nitrogen is first mineralised before being immobilised by microorganisms [5]. The nitrogen present in amino acids is mineralised when amino acids are further degraded into VFA and NH_3 [5]. Nowadays, it is known that microorganisms can also assimilate nitrogen directly from an N-containing substrate. In the Direct (DIR) hypothesis the nitrogen of substrates is always directly assimilated [5]. The parallel (PAR) scheme forms a bridge between the MIT and the DIR hypothesis, saying that part of the nitrogen is directly assimilited and the other part is mineralised first [5]. The PAR scheme is the most realistic representation of how microorganisms fulfil their nitrogen need.

2.3 Aerobic degradation of microbial residues

The first step in defining all SOM of microbial origin is the decay of biomass into necromass, which consists of several microbial polymers. These microbial polymers are then subject to extracellular enzymes. The released monomers will then be taken up by biomass, which is the last step in identifying new compounds and reactions in SOM degradation.

2.3.1 Biomass decay

Large differences exist in the exact polymer composition of necromass. A first important step to account for these differences is to distinguish between fungal and bacterial necromass [14].

Fungal necromass consists of four major polymers: chitin, glucan, proteins and melanins. Proteins have already been defined in Section 2.2.1. The three other polymer types are discussed

Table 5: Monomer characterisation of plant residues after oxidation

Monomer	**Chemical formula**	**-COOH**	**-Arom-OH**	**C=O**
Dihydroxybenzoic acid	$C_7H_6O_5$	1	2	0
3-hydroxy-4-methoxy-5-carboxybenzoic acid	$C_7H_8O_6$	2	1	0
2,3,4-trihydroxybenzoic acid	$C_7H_6O_5$	1	3	0

below. All newly introduced microbial monomers are summarised in Table 6. The chemical formulas of the microbial polymers are summarised in Table 7. The microbial polymers are further characterised by their carboxyl, phenolic and carbonyl functional groups in Table 8.

1. Chitin is the structural component of fungal cell walls [14]. $N-$Acetylglucosamines (NAG, $C_8H_{15}NO_6$) form the monomeric building blocks of chitin [8, 14]. Fungi also contain other polymers similar to chitin, such as chitosan (glucosamine monomers) and galactosamine polymers [14]. The NAG monomers are linked together with $\beta-(1-4)$ bonds by a dehydration reaction. As a result, the chemical formula of chitin is $(C_8H_{13}NO_5)_n$.
2. Glucan is a structural component of both the fungal cell wall and the fungal cell matrix [14]. It is a polysaccharide of glucose monomers, which are linked through glycosidic bonds by a dehydradation reaction. By consequence, glucan has a chemical formula equal to $(C_6H_{10}O_5)_n$.
3. Melanins are non-hydrolysable structures in the cell wall of fungi, which protect the cell wall against hydrolysis [14]. Melanins have a polymeric core, with monomers similar to phenols and quinones [14]. Attached to this core are proteins, carbohydrates and lipids [14]. Due to the lack of knowledge on its exact composition, melanin is not included in Table 7.

Necromass from bacterial origin consists of three major polymers: peptidoglycan, proteins and bacteran.

1. Peptidoglycan is a main compound of bacterial cell walls and consists of N-acetylglucosamine and N-acetylmuramic acid (NAM, $C_{11}H_{19}NO_8$) monomers [8, 14]. Peptidoglycan systematically alternates one NAG and one NAM monomer which are bound with $\beta-(1-4)$-glycosidic bonds by a dehydradation reaction. Additionally the NAM monomer has an oligopeptide side chain of approximately four amino acids [14]. The amino acids are bound through peptide bonds, which removes four H_2O molecules in total. The oligopeptide side chain is often cross-linked to another side chain, which would require the removal of one additional H_2O molecule. The last H_2O molecule removed is shared between two units and thus only counts as half. The overall composition of peptidoglycan therefore becomes that of one NAG, one NAM, four amino acids, minus six H_2O molecules (and often 6.5) removed for dehydration. This yields a chemical formula of $(C_{35}H_{54.28}N_{6.48}O_{16.8})_n$.
2. A number of bacteria have been shown to contain significant amounts of non-hydrolysable aliphatic biomolecules, called bacteran [14]. They derive from the condensation of complex lipids and are located in the cell wall [14]. Little information is available on their exact composition [14]. As a result, no general expression for the chemical formula is given.

To establish a stoichiometric equation for decay, we introduce the parameter $\%_{fungi}$, which represents the amount of fungal biomass at a given moment. The composition of fungal necromass can subsequently be characterised through the percentage of chitin ($\%_{chit}$), glucan ($\%_{gluc}$), protein ($\%_{prot,f}$)and melanin ($1-\%_{chit}-\%_{gluc}-\%_{prot,f}$) present in the necromass. By consequence, $1-\%_{fungi}$ is the amount of bacterial biomass at that same moment. Likewise, the composition of bacterial necromass is characterised through the percentage of peptidoglycan ($\%_{pept}$), protein ($\%_{prot,b}$) and bacteran ($1-\%_{pept}-\%_{prot,b}$). Values for the percentages could be estimated by using a description given by Throckmorton et al. (2012) [27], where necromass is characterised with pyrolysis-GC-MS. The characterisation is divided in extracellular polysaccharides, lipids, phenols, benzene, polysaccharide and N-compounds.

2.3.2 Depolymerisation of microbial residues

The microbial polymers need to be depolymerised before they can act as substrate for biomass growth themselves. From the five microbial polymers discussed, three can be hydrolyzed: chitin, glucan and peptidoglycan.

The bonds in chitin can be hydrolyzed by chitinase [7]. The glycosidic bonds in glucan can be hydrolyzed by glucosidase [7]. Peptidoglycan can be hydrolyzed by peptidoglycan hydrolase and several other enzymes depending on which bond is hydrolyzed. The stoichiometry of the three hydrolysis reactions is summarised in Table 9.

Little is known about the composition and decomposition of melanins and bacterans [14]. It has been shown that white-rot enzymes involved in lignin degradation can also degrade melanins [14].

2.3.3 Substrate uptake of microbial monomers

N-acetylglucosamine and N-acetylmuramic acid are the only monomers which are specific for MOM. The monomeric uptake of microbial monomers is the last step that needed to be defined from a chronological point of view.

2.4 Kinetics of aerobic degradation

In Section 2.2 and 2.3 the important components and reactions have been presented that occur during SOM degradation. To be able to establish a reaction network, additional information is necessary with regard to the kinetics of the above-mentioned reactions.

It is common to assume first order kinetics for depolymerisation reactions. This is in accordance with the traditional compartment models for SOM, such as RothC [4, 5], as well as with water treatment models such as the ASM3 model [5, 28]. The reaction rate r_S $[\frac{mol}{L \cdot h}]$ in first order kinetics follows Eq. 1:

Table 6: Monomer characterisation of microbial necromass

Monomer	**Chemical formula**	**-COOH**	**-Arom-OH**	**C=O**
N-acetylglucosamine	$C_8H_{15}NO_6$	0	0	1
N-acetylmuramic acid	$C_{11}H_{19}NO_8$	1	0	1

Table 7: Chemical formula of polymers in microbial necromass

Polymer	**Parameters**	**Chemical formula**
Chitin	-	$(C_8H_{13}NO_5)_n$
Glucan	-	$(C_6H_{10}O_5)_n$
Peptidoglycan	-	$(C_{35}H_{55.28}N_{6.48}O_{16.8}S_{0.16})_n$

$$r_S = \frac{dC_S}{dt} = k_{max} \cdot C_S \tag{1}$$

Where C_s is the substrate concentration $[\frac{mol}{L}]$ and k_{max} is the maximum rate parameter [1/h].

K_{max} is a measure of the recalcitrance of a polymer. Recalcitrance is the ability of a polymer to protect itself from extracellular enzymes through its intrinsic molecular characteristics [7]. A higher k_{max} leads to higher reaction rates and thus faster degradation. Finding correct value ranges for k_{max}, which only reflects recalcitrance of the molecule and no other stabilisation mechanisms is, however, difficult. While the characterisation of growth rates is regularly conducted under ambient temperature and pressure, the determination of depolymerisation is either focussed at industrial applications [29], with high temperatures and pressures, or at in-situ characterisation of depolymerisation in the soil. As a rule of thumb, it can be stated that polymers that require oxidative depolymerisation have a higher recalcitrance, and by consequence a lower k_{max} value [14]. Table 10 summarises the subdivision between polymers with higher and lower recalcitrance.

It is important to note that first order kinetics for depolymerisation are only valid when there is no limitation coming from another reactant, such as water or oxygen. To include the effect of substrate limitation, other than the one who has a strictly linear dependence, a substrate limitation term can be added to the equation for the reaction rate. The reaction rate is then calculated with Eq. 2 [5, 30]

$$r_{s1,s2} = k_{max_{s1,s2}} \cdot C_{s1} \cdot \frac{C_{s2}}{C_{s2} + K_{Inh}} \tag{2}$$

Where $r_{s1,s2}$ is the reaction rate for an equation with two substrates s1 and s2 $[\frac{mol}{L \cdot h}$, $k_{max_{s1,s2}}$ the maximum reaction rate $[\frac{1}{h}]$, C_{s1} the concentration of the substrate s1 which shows linear dependence $[\frac{mol}{L}]$, C_{s2} the concentration of the limiting substrate s2 $[\frac{mol}{L}]$ and K_{Inh} is the inhibition term $[\frac{mol}{L}]$ [6].

The metabolic reactions consist of substrate uptake reactions leading to microbial biomass growth. When characterizing growth reactions, we often use the specific growth rate μ instead of the standard reaction rate r_{growth}. The relationship between the two parameters is given by Eq. 3:

$$\mu = \frac{r_{growth}}{C_X} \tag{3}$$

Where μ is the specific growth rate $[\frac{1}{h}]$, r_{growth} is the reaction rate of the growth reaction $[\frac{mol}{L \cdot h}]$ and C_X is the biomass concentration $[\frac{mol}{L}]$.

Table 8: Functional groups of polymers in microbial necromass

Polymer	**Parameter**	**-COOH**	**-Arom-OH**	**C=O**
Chitin	-	0	0	1
Glucan	-	0	0	0
Peptidoglycan	-	1	0	6

Table 9: Depolymerisation of hydrolysable microbial polymers

Polymer	**C-compounds**	**H_2O**
Chitin	1NAC	-1
Glucan	1Glucose	-1
Peptidoglycan	1NAC + 1NAM + 4Amino Acids	-6

Table 10: Intrinsic recalcitrance of SOM polymers

High k_{max}	**Low k_{max}**
Cellulose	Lignin
Hemicellulose	Condensed tannin
Proteins	Melanin
Cutin	Bacteran
Chitin	
Peptidoglycan	
Glucan	

The specific growth rate μ is modelled with the Monod equation [6], given by Eq. 4:

$$\mu = \frac{dC_X}{C_X \cdot dt} = \mu_{max} \cdot \frac{C_S}{C_S + K_{Inh}} \tag{4}$$

Where μ is the specific growth rate $[\frac{1}{h}]$, μ_{max} is the maximum specific growth rate $[\frac{1}{h}]$ and C_S is the limiting carbon substrate $[\frac{mol}{h}]$.

The reaction rate is rewritten as Eq. 5 to show that the reaction rate depends linearly on the total biomass concentration:

$$r_{growth} = Y_{X/S} \cdot \frac{dC_S}{dt} = \mu \cdot C_X = \frac{-\mu_{max}}{Y_{X/S}} \cdot C_X \cdot \frac{C_s}{C_S + K_{Inhb}} \tag{5}$$

Where r_{growth} is the reaction rate of the growth reaction $[\frac{mol}{L \cdot h}]$, $Y_{X/S}$ is the growth yield of biomass on substrate S $[\frac{mol}{mol}]$, C_S is the substrate concentration $[\frac{mol}{L}]$, C_X is the biomass concentration $[\frac{mol}{h}]$ and μ_{max} is the maximum specific growth rate $[\frac{1}{h}]$.

For carbon substrates which do not contain nitrogen, the growth reaction also depends on the availability of NH_3. Additionally, also O_2 will participate in some growth reactions. Both dependencies can be modelled by adding a substrate limitation term to the reaction rate.

The decay reaction of biomass is also typically modelled as first order reaction [28].

Each reaction can potentially be inhibited by the concentration of a specific compound. Several inhibitions can be found in literature which influence degradation of soil organic matter. Tannins can bind with enzymes, which disables them and inhibits SOM degradation [18]. Xylose oligomers, released from hemicellulose hydrolysis, have been show to inhibit cellulose hydrolysis [29]. Low N-contents reduces the growth yield [31]. Onset of production of lignin-degrading enzymes from white-rot fungi might be linked to a secondary metabolism condition in response to a nutrient depletion [17]. The resulting monomers (or even oligomers) depend on the type of enzymes, and thus on the type of microorganisms available. Phenols also inhibit their own uptake by microorganisms. Several inhibition models are known to describe the dynamics of

microbial growth on phenol [32]. A commonly used model is the Andrews inhibition model, also known as Haldane model [32, 33], given by Eq. 6:

$$\mu_s = \mu_{max,s} \cdot \frac{C_s}{K_s + C_s + \frac{C_S^2}{K_i}} \tag{6}$$

Where μ_s is the specific growth rate on substrate s $[\frac{1}{h}]$, $\mu_{max,s}$ is the maximm specific growth rate on substrate s $[\frac{1}{h}]$, C_s is the substrate concentration $[\frac{mol}{L}]$, K_s is the half-saturation constant $[\frac{mol}{L}]$ and K_i is the inhibition constant $[\frac{mol}{L}]$.

2.5 Anaerobic degradation of soil organic matter

Anaerobic conditions can occur in certain soil types, e.g. peat soils [34], but also more generally speaking in subsoils. SOM degradation can still exist in anaerobic conditions, but involving pathways and SOM components different from those discussed in Sections 2.2 to 2.4.

In anaerobic conditions the microorganisms are mostly limited to (anaerobic) bacteria, as the vast majority of fungi require oxygen to exist. The anaerobic degradation of readily degradable, i.e. hydrolysable, SOM polymers is called anaerboic digestion [35]. Anaerobic digestion consists of a series of steps, each performed by a different fraction of the microbial community. The four major processes which make up anaerobic digestion are extracellular hydrolysis of polymers, acidogenesis, acetogenesis or anaerobic oxidation and methanogenesis. Each of the three last processes are performed by a different fraction of the microbial community. The hydrolysis converts the polymers into monomers such as monosaccharides, LCFA and amino acids [35]. Acidogenesis perfomed by acidogens then converts the simpler monomers into VFA. Acetogenesis by acetogens uses VFA as substrates for growth, which results in the production of acetic acid, CO_2 and hydrogen. Lastly, methanogens consume acetates and hydrogen during methanogenesis, while producing methane (CH_4) and CO_2.

A large fraction of recalcitrant polymers such as lignin and condensed tannin were thought to be undegradable in anaerobic conditions [34, 35]. The idea is now growing that even for these polymers microorganisms have adapted to degrade them in anaerobic conditions. Extensive research will be necessary to establish a second reaction network for anaerobic SOM degradation of all polymers, which is beyond the scope of this research.

3 Theoretical Framework

In this section, a reaction network for SOM degradation is established based on the results of the literature research described in Section 2. Choosing a finite set of compounds and reactions requires a large amount of simplifications. These simplifications are discussed and argued in this section. It has been chosen to only define an aerobic reaction network, as the anaerobic degradation still requires further documentation.

Choosing SOM polymer types Six plant polymers are included in the reaction network, i.e. cellulose, hemicellulose, proteins, cutin, lignin and condensed tannins. Starch has the same chemical formula and functional groups as cellulose, and is, therefore, not included separately. Due to limited knowledge on suberin's composition and degradation, it is not explicitly represented in the reaction network. Nonetheless, the aliphatic fraction of suberin can be accounted for through an equivalent concentration of cutin. Likewise, the aromatic fraction can be accounted for through an equivalent concentration of lignin. Hydrolysable tannins are not included in the reaction network either, because they are only present in plants in very low concentrations [14]. Although condensed tannins are also only present in low concentrations, they have been included due to their higher recalcitrance. The condensed tannins might therefore gain more importance throughout degradation due to their preservation. Additionally, five new microbial polymers are included as well, i.e. glucan, chitin, peptidoglycan, bacteran and melanin. Other glucosamine and galactosamine polymers can be accounted for through chitin due to their similar characteristics.

Choosing SOM monomer types Seven monomer types are represented in the reaction network, i.e. monosaccharides, monoaromatic compounds, amino acids, VFA, LCFA, NAM and NAG. Monosaccharides are crucial monomers for glucan, cellulose (or starch) and hemicellulose. Monoaromatic compounds are necessary to represent depolymerisation of lignin, condensed tannins and melanin. Amino acids are a product of depolymerisation of proteins and peptidoglycan. VFA are released during the depolymerisation of hemicellulose and lignin. Furthermore, amino acids can also react to amino acids, according to the MIT and PAR scheme. LCFA are the monomers of cutin and bacteran. LCFA are also used to represent lipids in the system, as LCFA are a type of lipid. NAM is released during peptidoglycan depolymerisation. Finally, NAG is a monomer both for chitin and peptidoglycan. In total eighteen SOM compound types are thus included in the reaction network. One global biomass component is also included. Additionally, some monomers defined previously can dissociate in water. As a result, dissociated forms of NAM, NAG, LCFA and VFA become part of the reaction network.

Choosing the non-organic compounds Subsequently, it is necessary to define the non-organic fraction of the compounds active in the degradation process. The crucial compounds for degradation are H_2O, O_2, NH_3 and CO_2. As mentioned in Section 2.2.2 H_2O_2 can also be used as oxidator for the depolymerisation of lignin and possibly other compounds. Nevertheless, the compound is not included in the reaction network. The oxidation reactions already contain high levels of uncertainty, and H_2O_2 chemistry is more complex than that of O_2. H_2O can dissociate to H^+ and OH^-. NH_3 can be protonated and form NH_4^+ in solution or be in equilibrium with NO_3^-. CO_2 is in equilibrium with H_2CO_3 and HCO_3^-. H^+, OH^-, H_2CO_3, HCO_3^-, NH_4^+ and NO_3^- also become part of the reaction network. The result of these decisions can be seen in Figure

3. This reaction network forms the framework for the implementation of SOM degradation in a numerical forward model.

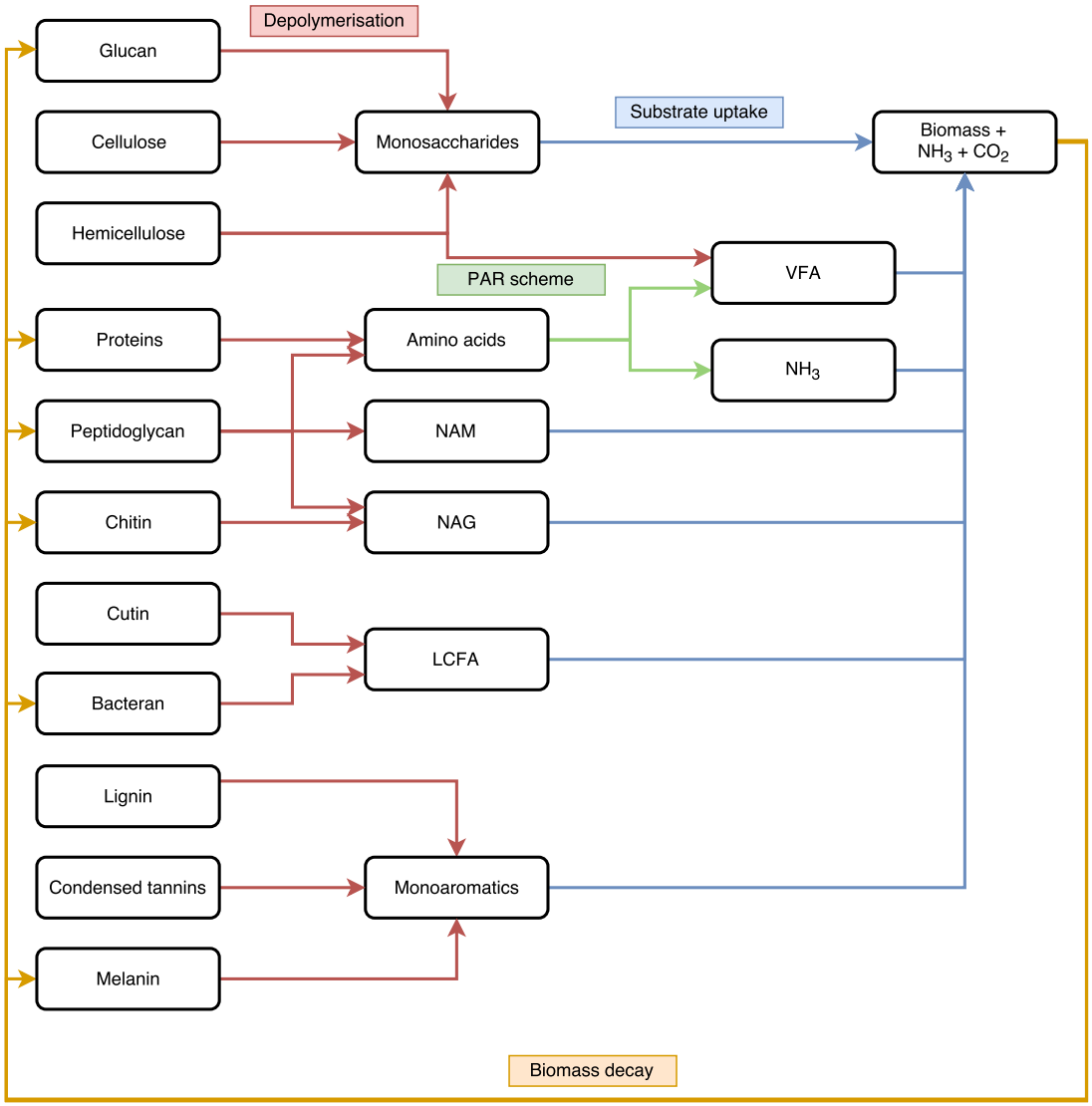


Figure 3: Complete reaction network

Characterisation of polymers To impose a mass balance for all reactions, it is necessary to characterise all compounds with a chemical formula. The polymers can be divided in three categories based on the way their chemical formula is set. In the first category, the polymers have a fixed chemical formula which closely represents their actual elemental composition. This category consists of cellulose, condensed tannins, proteins, glucan, chitin and peptidoglycan. In the second category, the chemical formula is calculated based on a few extra parameters that characterise the polymer. This category consists of hemicellulose ($\%_{C_6}$, η_{Ac}), lignin ($\%_G$, $\%_S$, $\%_{ArO}$) and cutin ($\%_{C_{18}}$) through additional parameters. The third category is that of polymers

whose composition is ill-defined and therefore the chemical formula and other characteristics are solely a rough estimation of those characteristics. Melanin for example is known to have aromatic compounds. In this reaction network it is thus represented through the same chemical formula and characteristics as lignin. Bacteran is known to contain LCFA monomers. It is thus represented through the same characteristics as cutin.

Characterisation of monomers The SOM compound types for monomers need to compass a larger variety of compounds through only one model compound. Therefore, we select a model compound whose characteristics are closest to the average of those characteristics of the other compounds it has to represent. For most compound types, this model compound corresponds with an actual, perfectly defined compound. By consequence, all other characteristics besides chemical formula, can also be deducted directly from this compound. The model compound for monosaccharides is glucose, because both glucan and cellulose solely consist of hexose monomers. For hemicellulose this can also be the case, though pentose is more common. NAM and NAG are already actual compounds and therefore they do not require the choice of a model compound. Acetic acid is the model compound for VFA, for several reasons. Acetylated hemicellulose releases acetic acid, while amino acid and lignin might also react to other VFA. Furthermore, acetic acid has the shortest chain length of all VFA. This makes acetic acid suitable to solve discrepancies in the carbon mass balance. The model compound for monoaromatic compound types is dihydroxybenzoic acid. Dihydroxybenzoic acid is the oxidative degradation product closest to the average of the other products represented in Table 5. Furthermore, dihydroxybenzoic acid is similar (or identical) to some breakdown products of both hydrolysable and condensed tannins [18]. The characterisation of the LCFA compound type forms a minor exception, as the characterisation is only indirectly conducted by choosing a perfectly defined and actual model compound. Instead, it is achieved through the weighted average of the model compound for C_{16} LCFA and that for C_{18} LCFA. The chosen model compound for C_{16} is 10,16-dihydroxypalmitic acid and that for C_{18} is 18-hydroxyoleic acid. The weighted average is achieved through the percentage of C_{18} ($\%_{C_{18}}$) in cutin. The last adjustment regarding chemical formulas is that sulphur (S) is not included in the reaction network, because it is not an essential component of degradation and present in lower concentration in proteins and amino acids when compared to C, H, O and N.

Setting the reaction stoichiometry **Include matlab code in appendix** Once the reactions, compounds and chemical formulas have been set, the stoichiometric equation can be defined. Generally speaking, the stoichiometric equation is set by defining the coefficient of the SOM reactants and products which ought to cover the C-balance, and then closing the mass balance for N and O with NH_3 and H_2O. Oxidation reactions additionally need to be balanced for H by H^+, then balanced for charge by e^-. Subsequently, this reaction is balanced for electrons with the half-reaction of O_2 reduction. For reduction reactions, the reaction is balanced with the half-reaction of the oxidation of H_2O. For metabolic reactions, this scheme is followed for both the catabolic and the anabolic reaction, through the growth yield. The main challenge in setting the stoichiometric equation is dealing with discrepancies between the chemical formula of the actual compounds and that of the model compound. For hemicellulose, the actual monomer can be pentose ($C_5H_{10}O_5$), but the model monosaccharide has a chemical formula set to be $C_6H_{12}O_6$. The stoichiometric coefficient for the monosaccharide is thus changed to 5/6 instead of one, to maintain a closed carbon mass balance. For lignin oxidation, a discrepancy exists between the number of carbon atoms (NoC) of the polymer and the NoC of the monomer, i.e.

dihydroxybenzoic acid. This is dealt with by accounting for the extra C atoms through VFA production. Finally, the monoaromatic compound only consists of one aromatic ring ($NoC = 7$), while catechin is made out of two rings ($NoC = 15$). This is accounted for by changing the stoichiometric coefficient of the monoaromatic compound to 15/7, as to add 15 carbon atoms to the monoaromatic compound concentration per reaction.

Defining the reaction kinetics A reaction is defined by both its stoichiometry and its kinetics. As discussed in Section 2.4, the depolymerisation and decay reactions are modelled through first order kinetics, with their own concentration as reaction driving concentration. Additionally, substrate limitation can be accounted for when other limiting substrates are used. Values for the maximum rate parameter have to be chosen from experimental data. Substrate uptake and the related biomass growth are assumed to follow Monod kinetics. Maximum growth rates too, can be defined based on experimental studies, but, moreover, they can be estimated through a method developed by Kleerebezem et al. (2010) [25], pertaining on thermodynamic principles. The maximum growth rate is then calculated from the ΔG_f˚at 25˚C of all reactants and products, from the stoichiometry of the catabolic and anabolic (half-)reactions and from the growth yield. For substrates with a NoC equal to or larger than seven, the growth yield then still needs to be set through experimental data. For the substrates with NoC equal to or lower than six, the growth yield can be estimated by the same method, using only the number of carbon (NoC) atoms of the substrate and the oxidation state γ of the substrate [25].

Choosing inhibitions Another aspect of kinetics in the reaction network are inhibitions. Only the inhibition caused by phenols, i.e. substrate limitation by phenols [33], is included. This decision is made, because monoaromatic compounds are expected to have an important effect on the bulk properties of the SOM mixture. The reason is that monoaromatic compounds are the only dissolved aromatic (and phenolic) compounds. Not including the inhibition would underestimate the amount of monoaromatic compounds in the SOM mixture. Furthermore, the inhibition is well described in literature. Moreover, the inclusion of this inhibition could be used as a framework to add additional inhibitions in the future. The last issue concerning the kinetics is the degradation of amino acids to VFA and NH_3. In the DIR scheme this reaction rate would be set to zero. In the MIT scheme the uptake of amino acids would be set to zero and a maximum rate parameter for the reaction to VFA and NH_3 would need to be defined. In this reaction network we define a flexible PAR scheme, where the reaction rate for the reaction to VFA is linked to the uptake rate for amino acid uptake. Any intermediate situation between MIT and DIR can therefore be modelled. A new variable is introduced, which we will call the amino acid assimilation efficiency η_{AsAA}. The η_{AsAA} represent the percentage of amino acids that are directly assimilated by microorganisms, without first degrading to VFA. The DIR scheme corresponds with a η_{AsAA} value of one and the MIT scheme with a value of zero. Any values in between correspond with PAR schemes, which lean more towards the MIT or to the DIR scheme.

4 Toolbox

The reaction network defined in Section 3 is implemented as a forward numerical model in MATLAB, based on the gray modelling toolbox of Van Turnhout et al. (2016) [6]. Section 4.1 lists all input variables necessary to run the forward model. The forward model itself is described in Section 4.2. The processing of the output from the forward model is discussed in Section 4.3. A schematic representation of the forward model is given in Figure 4. At last, Section 4.4 presents the implementation of the DREAM algorithm from Vrugt et al. (2016) [36], which is used for a Bayesian Inference analyse, described in Section 5.3.

4.1 Input variables for forward model

The forward model requires input variables, which can be divided in four categories. All variables are summarised by category in Table 11. The first category contains variables to characterise the SOM compound types. These parameters can be measured prior to running the forward model and have already been discussed in Sections 2 and 3. The second category consists of the variable pertaining to the initial conditions, i.e. the concentrations of all compounds j that are initially (t_0) present $C_{j,0}$, the batch temperature T [K], the volume of the liquid phase V_l [l], the volume of the gas phase V_g [l] and the batch pressure p [atm]. The third category encloses input variables which cannot be determined prior to running the forward model, due to high uncertainty. Rather than implementing the forward model with fixed values, the parameters become variables. Doing so allows us to easily change the values after each run. This feature is used to apply Bayesian Inference on these parameters (see later). The variables present in the third category are all the k_{max} for the depolymerisation reactions, the growth rates (optional), yields (optional for monomers with $NoC > 6$), the k_{max} for decay, K_s and K_i as inhibition parameters for phenol substrate limitation. The fourth and last category contains variables which the ODE solver uses to solve the above set of differential (reaction) equations. These variables are the initial time t_0, the end time t_{end} and the time step Δt.

Table 11: List of input variables

Type of input	**Parameters**
Characterisation	$\%_{C_6}$, η_{Ac}, $\%_G$, $\%_S$, $\%_{ArO}$, $\%_{C_{18}}$, $\%_{BiPh}$ $\%_{fungi}$, $\%_{chit}$, $\%_{gluc}$, $\%_{prot,f}$, $\%_{pept}$, $\%_{prot,b}$
Initial conditions	$C_{j,ini}$ $\forall$ compounds j present at t_{ini} T, p, V_l, V_g
Large uncertainty	$k_{max,j}$ $\forall$ depolymerisation reactions j $\mu_{max,j}$ optional for substrate uptake reactions $k_{max,decay}$ K_s and K_i for phenol inhibition $Y_{X/S}$ optional for substrates with NoC $\leqslant 6$ η_{AsAA} for amino acid reaction rates
Solver	t_{ini}, t_{end}, Δt

4.2 Forward model

The forward model consist of a static and a dynamic component. The static component initialises a system of differential equations, which represent the reactions that are active throughout aerobic degradation of SOM. The initialisation is conducted in three stages: 1. initializing the compounds with their characteristics, 2. initialising reaction stoichiometry and reaction rates of kinetic reactions and 3. initialising the equilibrium reactions involving the liquid and gas phase. The dynamic component solves the system of differential equations defined in the previous stages.

Initialisation of compounds Initialising the compounds present in the reaction network is achieved by creating a cell array which saves their names, chemical formula, amount of functional groups, gibbs free energy and oxidation state. Another structure saves the index of each compound within the array. The chemical formula of the polymers are defined by first creating the monomers as compounds. The chemical formula of the monomers is always hard-coded, following the description in Table 1 and Table 6. When multiple monomers are present in one polymer, the formulas are combined together according to the composition as described in Table 2 and Table 7. For hemicellulose the formula of acetic acid is multiplied by factor η_{Ac} and then added to the formula for the monosaccharide. For cutin the formula of the C_{18} LCFA is multiplied by factor $\%_{C18}$ and then added to the formula of the C_{16} LCFA. For lignin the formulas of the G, S and P unit are respectively multiplied with factors $\%_G$, $\%_S$ and $(1-\%_G-\%_S)$ and then added together. For polymers with multiple hydrolysable bonds the formula of H_2O is then multiplied by the amount of hydrolysable bonds (see Table 4 and Table 9) and subtracted from the monomeric formula. For lignin and tannin two hydrogen atoms are removed from the monomeric formula. The formula for melanin is set equal to the formula of lignin. The formula for bacteran is set equal to the formula for cutin. Subsequently, the amount of functional groups (carboxyl, phenolic and carbonyl) are hard-coded for each compound, following Table 3 and Table 8. Only the amount of phenolic groups in lignin varies and is equal to $\%_{BiPh}$. The $\Delta G_f^\circ$of each compound is hard-coded in the model as well. All values are shown in Table 12. The last characteristic is the oxidation state of carbon atoms in the SOM compounds. The oxidation state γ is calculated following Eq. 7

$$\gamma = \frac{-1 \cdot NoH + 3 \cdot NoN + 2 \cdot NoO}{NoC} \tag{7}$$

Where NoH is the number of hydrogen atoms in the compounds, NoN the nitrogen atoms, NoO the oxygen atoms and NoC the carbon atoms.

Initialisation of stoichiometric equations The initialisation of the kinetic reaction consists of setting both the stoichiometry and the kinetics of the reaction. For the non-metabolic reactions, setting the stoichiometry means closing the mass and electron balances. This is achieved through a series of subsequent steps. The first step is defining the stoichiometric coefficients of the carbon-containing SOM compounds of each reaction. These coefficients are based on the monomer composition, which are either hard-coded, or function of additional parameters. This has been discussed in Section 3. Based on the requirement of a closed mass balance for O, N and H, and a charge balance, the stoichiometric coefficients of H_2O, NH_3, H^+ and e^- are then automatically calculated. For redox reactions, i.e. the reactions where the coefficient for e^- is non-zero, the equation is balanced with either the reduction of O_2 to H_2O or the oxidation of H_2O to O_2. For the substrate uptake reactions, the same process is followed for both the anabolic and the catabolic reaction. The two reactions are then balanced by $Y_{X/S}$. For the substrates

with NoC $\geqslant 7$, the $Y_{X/S}$ has been defined as an input variable. For the other substrates, the forward model first checks if the parameter has been defined as an input variable. If not, it is calculated with the method of Kleerebezem et al. (2010) [25] based on thermodynamic principles. The method uses the ΔG_f° of all reactants and products in the anabolic and catabolic reactions, and the stoichiometric coefficients of these two reactions. Once all stoichiometric equations have been calculated, the coefficients of H_2O and O_2 are overwritten to be set to zero. By setting the consumption and production of H_2O to zero, the volume of the liquid phase stays constant. By setting the consumption and production of O_2 to zero, it is assumed the environment is not limited by O_2.

Initialisation of kinetics To set the kinetics of the reaction, we define the reaction driving concentration $C_{driving}$, the maximum rate parameter k_{max} or μ_{max} and the inhibitions, including Monod kinetics. For depolymerisation reactions, the $C_{driving}$ is the polymer concentration. For substrate uptake reactions and for biomass decay, the $C_{driving}$ is the biomass concentration C_X. The k_{max} for decay and depolymerisation reaction has to be passed to the forward model as an input variable. For the μ_{max} the forward model always checks if the parameter has been defined as an input variable. If not, it is calculated by a method from Heijnen et al. (2010) [37], based on thermodynamic principles. The rate of the degradation of amino acids to VFA and NH_3 is calculated by linking the input variable η_{AsAA} with the μ_{max} of amino acid uptake, by following Eq. 8:

$$k_{max,AA \Rightarrow VFA} = 2 \cdot (1 - \eta_{AsAA}) \cdot \mu_{max,AA} \tag{8}$$

Where $k_{max,AA}$ is the k_{max} for the reaction of amino acids to VFA $[\frac{1}{h}]$, $\mu_{max,AA}$ is the theoretical μ_{max} with amino acids as substrates $[\frac{1}{h}]$ and η_{AsAA} is the efficiency of assimilation of amino acids $[-]$.

We assume that $\mu_{max,AA}$ as defined so far is the one valid when half the amino acids react to VFA and half are taken up, i.e. corresponding with η_{AsAA} equal to 0.5. In other cases, the actual $\mu_{max,AA}$ is updated from Eq. 9:

$$\mu_{max,AA,new} = 2 \cdot \eta_{AsAA} \cdot \mu_{max,AA,old} \tag{9}$$

Each inhibition is defined by the type of inhibition, the reaction it is acting upon and the inhibition parameters. There are six monod kinetics inhibitions defined for the six types of substrate uptake reactions, aside from monoaromatic uptake. Furthermore there are four substrate limitation inhibitions by NH_3 on the substrate uptake of monosaccharides, VFA, monoaromatic compounds and LCFA. The value of K_{inh} is set at 0.001 for these ten inhibitions. At last, there is the inhibition of phenol on its own uptake, specified with K_S and K_i which have been passed to the forward model as input variables. The monod kinetics are already present in the toolbox of Van Turnhout et al. (2016) [6]. The Haldane kinetics for phenol substrate limitation is added.

Initialisation of equilibrium reactions The above-mentioned reactions are all kinetic reaction. In addition, equilibrium reactions occur in the system. In the liquid phase, certain compounds can dissociate. Initialising the dissociation reactions follows the same approach as Van Turnhout et al. (2016) [6]. We indicate that H^+, OH^-, H_2O, $C_2H_4O_2$, $C_2H_3O^-$, H_2CO_3, HCO_3^-, NH_3, NH_4^+ and NO_3^- are all present in the system. The dissociation constants of these compounds are available in Orchestra's Minteq database. Additionally, the dissociation of the

new monomers is also defined. This consists of the dihydroxybenzoic acids, NAM and LCFA. For the monomaromatics, the pKa of the 2,4-dihydroxybenzoic acid is used, because their pKa values form an intermediate value between the pKa of the several isomers. The pKa of N-acetyl muramic acid is estimated through the pKa of the carboxyl group of glycolic acid. For the long chain fatty acids, the estimated value of pKa for a palmitic acid is used [38]. The values can be found in Table 12. The gas phase consists of O_2, which is also the filler gas, CO_2 and NH_3. The concentration of CO_2 in the gas phase is in equilibrium with the concentration of H_2CO_3 in solution. The concentration of NH_3 is in equilibrium with the concentration of NH_4^+ in solution. The equilibrium is governed by Henry's law [6]. The Henry constant needs to be defined.

An overview of all parameter values which are hard-coded in the forward model is given in Table 12.

Table 12: List of hard-coded parameter values

Parameter category	**Parameters**	**Values**
ΔG_f°[kJ/mol]	Monosaccharide (glucose)	-917 [39]
	VFA (acetic acid)	-396.41 [39]
	Amino Acid	-500 ([-700;-200] [39])
	Dihydroxybenzoic acid	-551.8 [40]
	16-hydroxy-palmitic acid	-1332.25 [41]
	N-acetyl-glucosamine	-855.23 [42]
	N-acetylmuramic acid	-1153.68 [43]
Logarithmic acid dissociation constant (pK_a)	For inorganic fraction	Minteq Database
	Dihydroxybenzoic acid carboxyl	4.48
	phenolic group	8.83
	phenolic group	12.6
	N-acetyl muramic acid carboxyl group	3.83 (est)
	Palmitic acid carboxyl group	4.75 (est)
Inhibition constants K_{inh} [mol/l]	For NH_3 substrate limitation	0.001
	For monomeric substrates (monod)	0.001

Solving the reaction network Solving the system of equations is done by an ODE solver and Orchestra. The code for this was directly taken from Van Turnhout et al. (2016) [6]. The main output of the forward model is an array containing the concentrations of all compounds at each time step between t_0 and t_{end}, accompanied by the array containing all compounds and their characteristics, i.e. name, index, chemical formula, γ, $\Delta G_f^\circ$and functional groups.

4.3 Output processing

The final step is processing the output to generate meaningful results. The most basic results are plots of the concentration in time for all compounds. Additional plots are then generated by combining those concentrations with chemical properties of the corresponding compound and summing these results for relevant compound combinations. These plots demonstrate the dynamic nature of the bulk properties of SOM. All plots are listed below:

1. The concentration, expressed in Cmol, in time of the total soil organic carbon (SOC), by

multiplying the concentration of all SOM compounds with the coefficient for C in their chemical formula and then summing all the products.

2. The concentration in time of SOC associated with aromatic compounds, by multiplying the concentration of the aromatic compounds, i.e. lignin, tannin, monoaromatic compounds and melanin, with the coefficient for C in their chemical formula and then summing all the products.

3. The concentration in time of SOC associated with aliphatic compounds, by subtracting the SOC in aromatic compounds from the total SOC.

4. The average oxidation state in time of C in all SOM compounds, by multiplying all concentrations, expressed in Cmol, with the corresponding oxidation state, summing these products and then dividing by the total SOC.

5. The ratio in time between the aliphatic and the aromatic SOC in time, by dividing the concentration of SOC in aliphatic fraction by the concentration in the aromatic fraction.

6. The ratio in time between the concentration of SOC and nitrogen, by dividing the total SOC by the concentration of nitrogen both in N-containing compounds, i.e. proteins, amino acids, chitin and peptidoglycan, and in NH_3.

7. The concentration, expressed in mol, in time of carbonyl, carboxyl and phenolic groups, by multiplying the concentration of each compound, in mol, by the respective amount of

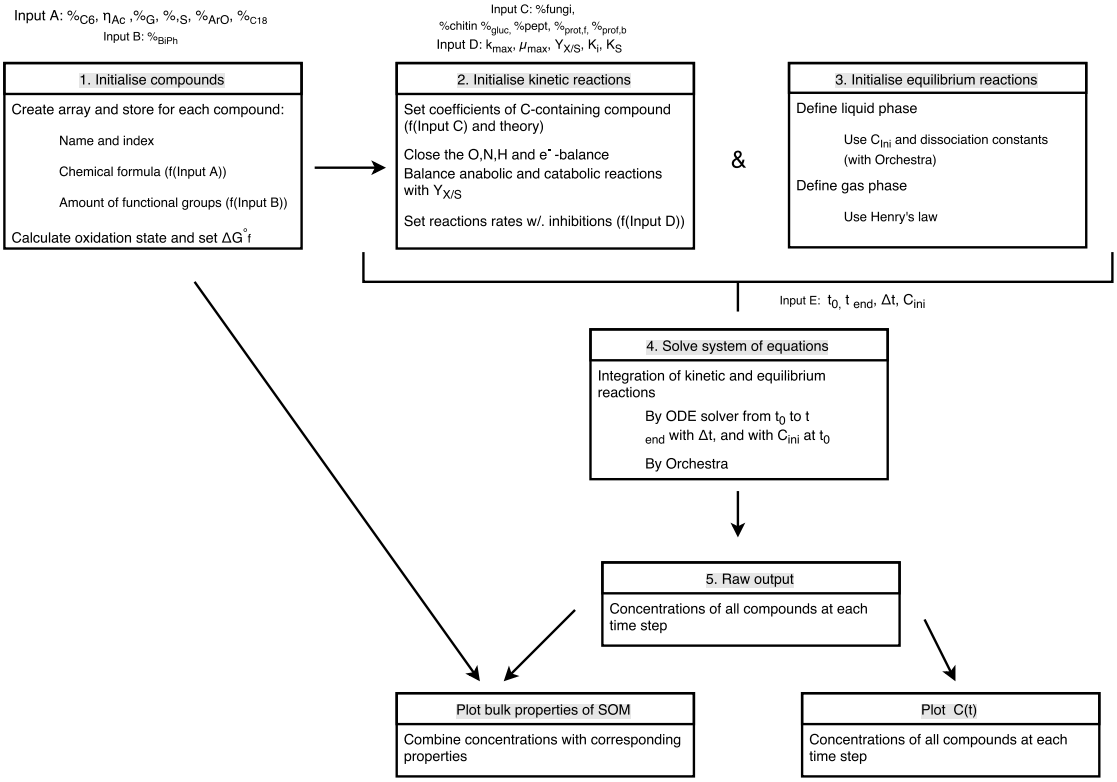


Figure 4: Schematic representation of the forward model

the functional group.

4.4 DREAM and hypothetical dataset

To be able to apply Bayesian Inference analyses on the forward model, the DREAM algorithm of Vrugt et al. (2016) [36] is included in the toolbox. The implementation is adopted from Van Turnhout et al. (2016) [6]. As an input, the DREAM module requires a dataset, a forward model, a number of iterations and a set of parameters on which the analysis is applied, including realistic value ranges. The output of the forward model is altered by prior processing, to mimic the measurement points of the dataset.

The DREAM module is expanded with a script that creates a hypothetical dataset. First, the forward model is run with specified values for all input variables, including the uncertain parameters. Second, the output variables of the forward run are processed to model the variables of interest, i.e. the measured variables in a real dataset. Third, we select only the values at certain time intervals, corresponding to the sample times in a real dataset. Lastly, noise is added to the output, to mimic the occurrence of measurement errors in a real dataset. The noise is normally distributed and defined by a standard deviation. Where applicable, the output is truncated to avoid negative concentrations after noise addition.

5 Method

With the implementation finalised, the toolbox can now be deployed as an analysis tool. We conduct three analyses. First, we initialise and run the toolbox for one example situation, to yield predictions on SOM and its properties. Second, we set up an analysis to choose an experimental design for quantification of uncertain parameter values. The third and last analysis illustrates the comparison between two reaction networks when the user of the toolbox is only interested in describing a specific dataset.

5.1 Forward run with baseline values

The output of the forward model can be interpreted as predictions on SOM degradation for the specific situation described by the values of the input variables. In this section we present one such situation, whose predictions are then addressed and discussed in Sections 6 and 7. The input variables are divided into three categories. In the first category are the uncertain parameters which cannot be measured directly (referred to as difficult to measure) and which are not bound to a specific situation. The suggested values for these parameters will be referred to as baseline (i.e. standard) values, which users can use without conducting further research. In the second category are variables that can be quantified by measurements prior to running the analysis or easily estimated based on the origin of the plant OM and data in literature. The third and last category contains the input variables that can simply be imposed on the system, such as temperature or the initial concentration. For illustrative purpose, we limit the presence of SOM compounds of plant origin to lignin and cellulose.

1. Difficult to measure input variables Not surprisingly, the input variables that are difficult to measure are also the variables with the most uncertain value. Due to the limited experimental data available, the values merely represent rough estimations. For the k_{max} parameters, we simply make a distinction between low and high recalcitrance, as done in Table 10. We assume that the slow reaction rates are approximately a factor 100 smaller than the fast reaction rates. This factor can be observed when comparing the reaction rates in in-situ measurements [44]. The measured values themselves are lower, resp. 0.06 1/d and 0.0005 1/d, but this can be attributed to the effect of stabilisation mechanisms. We arbitrarily decide to set the values to be a factor 5 faster than the in-situ parameters. Furthermore, we assume the growth rates on LCFA and monoaromatics to be as slow as the fast k_{max} of easily degradable polymers. The decay rate is set to be 0.8 1/d, which is slower than the growth rates of standard substrates (e.g. glucose), but faster the depolymerisation rates. The yields for LCFA and monoaromatic compounds is set to lower than the general heterotrophic yield, i.e. approximately 0.50 $\frac{cmol}{cmol}$. Next, we specify the growth yield for all other monomers with NoC $\leqslant 7$ to be 0.65 cmol/cmol, i.e. the general value for heterotrophic growth yield in ASM3 [28, 45]. It should be noted that the growth yield in the forward model is defined on a mol/mol basis, rather than cmol/cmol. By consequence, 0.65 needs to be multiplied with -NoC. The $\eta_{As,AA}$ is set to 0.5 to model an intermediate scenario between the DIR and MIT scheme. Table 14 summarises all values of the input variables. The chosen values are shown in Table 13, together with a hypothetical range within which the real value could fall.

Table 13: Baseline values for uncertain input variables

#	Parameter	Baseline value	Hypothetical range
1	$k_{hyd,cellulose}$	0.3 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
2	$k_{ox,lignin}$	0.003 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
3	$k_{hyd,protein}$	0.3 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
4	$k_{hyd,chitin}$	0.3 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
5	$k_{hyd,glucan}$	0.3 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
6	$k_{hyd,peptidoglycan}$	0.3 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
7	$k_{ox,bacteran}$	0.003 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
8	$k_{ox,melanin}$	0.003 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
9	$\mu_{monoaromatic}$	0.3 1/d	$[0.1 \cdot \mu_{hyp}; 10 \cdot \mu_{hyp}]$
10	μ_{LCFA}	0.3 1/d	$[0.1 \cdot \mu_{hyp}; 10 \cdot \mu_{hyp}]$
11	k_{decay}	0.8 1/d	$[0.1 \cdot k_{hyp}; 10 \cdot k_{hyp}]$
12	$K_{s,monoaromatic}$	0.01 1/d	$[0.005; 1]$
13	$K_{i,monoaromatic}$	0.6 1/d	$[0.01; 1]$
14	Y_{LCFA}	- 8 mol/mol	$[-9; -7]$
15	$Y_{monoaromatic}$	-3.5 mol/mol	$[-4.2; -1.4]$
16	$Y_{X/S}$ for NoC $\geqslant 7$	$-0.65 \cdot NoC$ mol/mol	$[-0.55 \cdot NoC; 0.75 \cdot NoC]$
17	η_{AsAA}	0.5	[0;1]

2. Measurable input variables The measurable input variables relate to the composition of the initial OM, in this case cellulose and lignin. The composition of cellulose is fixed. On the other hand, lignin is characterised with regard to its monomeric units and linkage. In this case, we represent lignin as having an equal amount of each unit ($\%_G = \%_S = \%_P = 1/3$), which is typical for lignin of grasses [14]. Furthermore, we specify that half of the linkages are aryl bonds ($\%_{ArO} = 0.5$) and 20% are biphenyl bonds ($\%_{BiPh} = 0.2$), which falls within a realistic distribution of bond types [14]. The presence of biomass implies the gradual supply of microbial residues through decaying biomass. We assume that half the necromass has fungal origin ($\%_{fungi} = 0.5$). A rough estimate of the composition of fungal necromass is conceived based on Throckmorton et al. (2012) [27]. The amount of chitin is set to 30% ($\%_{chit} = 0.3$), that of glucan to 45% ($\%_{gluc} = 0.45$) and that of proteins to 10% ($\%_{prot,f} = 0.1$). Similarly, the composition of bacterial necromass has to be specified. As such, we define the amount of peptidoglycan to be 75% ($\%_{pept} = 0.75$), that of bacterial protein to be 10% ($\%_{prot,b} = 0.1$). Although cutin is absent as plant residue, we still need to define a hypothetical percentage of C18 (vs. C16) LCFA. This parameter will be used to quantify the bacteran that enters the system through biomass decay. We arbitrarily set the amount of C18 LCFA to be 50% ($\%_{C18} = 0.5$).

3. Imposed input variables The hard-coded reaction environment currently associated with the forward model is a closed and perfectly mixed batch, with gas venting when total gas pressure p increases above the initial pressure p_0 [6]. The user can impose the batch temperature T, the initial pressure p_0, the gas volume V_g and the liquid volume of H_2O, V_l. In this case, the batch has a liquid volume V_l of one liter H_2O and equivalent gas volume V_g. The batch temperature T is specified as 283K (or approximately 10 °C) to mimic a realistic soil water temperature. The p_0 is set (and maintained) at 1 atm. The batch specifications are summarised in Table 15. The remaining imposed input variables are the initial concentrations. The cellulose concentration is chosen to be approximately 200 g/l, based on other published experimental designs where the weight/volume percentage concentration (w/v) of cellulose ranged from 5% to 20% [46], i.e. a

concentration [g/l] of 50 to 200 g/l. In both experimental designs of this section, the lignin concentration [g/l] is close to fifty percent of that of cellulose. This lignin-to-cellulose ratio resembles the lignin-to-cellulose ratio in leaf litter [14]. NH_3 is added to obtain a C/N ratio of 30, which is a realistic C/N ratio for fresh litter input [47]. Acetic acid and dihydroxybenzoic acid are included in low concentrations to improve the stability of the initial Orchestra calculations. Given that the model uses concentration in units of mol/l, it is required to convert the value. Therefore, the concentration in g/l, the molecular weight which is based on the chemical formula, and the concentration in mol/l are all included in Table 16.

With the above-mentioned values of input parameter, the reaction system is fully initialised. We decide to simulate the behaviour of SOM over a period of thirty days.

5.2 Evaluation of dynamic SOM characterisation

During the course of the simulation, the bulk properties of SOM will change due to the disappearance of some compounds and creation of others, which have varying properties. The exact bulk property depends on the concentration of each compound and on the properties of each individual compound. The course of the concentrations rely on highly uncertain kinetic parameters. For this reason, the validity of the result interpretation would be highly uncertain as well. However, there is one aspect of the simulation which has already been well defined, i.e. the individual compounds and their characteristics. The interpretation of the simulated results will thus focus on the range of values covered by the bulk property. For any simulated bulk property, a maximum and minimum value can be defined beforehand. The maximum value is the highest value of the property when looking at individual compounds. Likewise, the minimum value will be the lowest. By consequence, the simulated range of the bulk property will lie within these

Table 14: Values of input variables

Type of input	**Parameter**	**Value**
Characterisation	$\%_G$	1/3
	$\%_S$	1/3
	$\%_{ArO}$	0.5
	$\%_{BiPh}$	0.2
	$\%_{fungi}$	0.5
	$\%_{chit}$	0.30
	$\%_{gluc}$	0.45
	$\%_{prot,f}$	0.10
	$\%_{pept}$	0.75
	$\%_{prot,b}$	0.10
	$\%_{C18}$	0.5

Table 15: Values of imposed batch specifications

Parameter	**Value**
T [K]	283
p [atm]	1
V_l [l]	1
V_g [l]	1

Table 16: Initial concentrations

Compound	**C** $[\frac{g}{l}]$	**MW** $[\frac{g}{mol}]$	**C** $[\frac{mol}{l}]$
Cellulose	200.88	162	1.24
Lignin	93.8	187.6	0.5
Biomass C_x	0.0452	22.6	0.002
NH_3	6.8	17	0.4
Acetic acid	0.006	60	0.0001
Dihydroxybenzoic acid	0.3293	164.65	0.002

two limits and can possibly be significantly narrower. To evaluate the validity of these results, we will compare them to experimental data of humic substances.

Hypothesis The bulk properties of 'humic' substances should fall within the theoretical range of properties the model can offer. We hypothesis this result, because the identification of several compound types and their properties is based on a rigourus literature review. If the properties of humic substances lie beyond the range offered by this model, this either contradicts the statement that humic substances are actually organic compounds of plant and microbial origin, or that an important compound is missing in our model. Furthermore, we expect that the range of values for properties covered by the simulation is narrower than the full potential range of values.

Mode of evaluation To evaluate the ability of the simulations to characterise actual SOM mixtures, we compare the results of the simulation with properties of model humic molecules. Model humic molecules are molecules that are expected to be observed in SOM samples, but hypothetical SOM molecules whose properties are averages of experimentally determined properties [48]. We use two such molecules in this analysis, i.e. the Temple-Northeastern-Birmingham (TNB) molecule and the Schulten SOM molecule [48–50]. The characteristics of both model molecules are given in Table 17.

Table 17: Characterisation of TNB and Schulten SOM

Model	**Chemical formula**	**Ox**$_{C,avg}$	$\%_{COOH,avg}$	$\%_{AromOH}$
TNB	$C_{38}H_{39}O_{16}N_2$ [49]	-0.0263 (calc)	7,89 [49]	5.26 [49]
Schulten SOM	$C_{342}H_{388}O_{124}N_{12}$ [50]	-0.3041 (calc)	8.56 [50]	2.72 [50]

5.3 Experimental design selection for improvement of baseline values

The analysis introduced in this section brings forth the initial step in establishing a more adequate set of baseline values than the one suggested in Section 5.1, i.e. designing an experiment to quantify the parameters of interest.

After gathering data from any actual experiment, DREAM can be deployed to assess the identifiability of certain parameters of a postulated reaction network based on the measurements. When the identifiability of a certain parameters is low, it is then possible to add or change measurements obtained in an ensuing experiment. This is an iterative process. Due to the large

cost and time investment of conducting such experiments, we aim to find an experimental design which already provides an excellent starting point.

To achieve this, we need to realise that poor identifiability of a parameter arises from several factors. A first source of poor identifiability is that none of the monitored variables is sufficiently influenced by, i.e. not sensitive for, the parameter in question. Another factor is that the parameter in question does influence one or more monitored variables, but that it is correlated to another parameter. Due to the correlation, it is not possible to identify those parameters separately. The last reason is that the postulated reaction network is insufficiently correct and/or complex to explain the measured data. The poor identifiability which follows from insensitivity and correlations not captured in the presumed network cannot easily be evaluated prior to the actual experiment. On the other hand, poor identifiability arising from insensitivity and correlations which is inherent to the structure of the reaction network can and should be avoided.

For this purpose, we apply DREAM on two hypothetical datasets A and B to conduct what is essentially a sensitivity analysis. Both hypothetical datasets are created as explained in Section 4.4, representing two experimental designs A and B. This is the type of dataset that would be observed during a real experiment if the reaction network presented in this paper was an exact representation of the actual SOM degradation. In our case, the error between the simulated data and the measured data can now be entirely explained by measurement errors. The identifiability of the parameters of interest are compared for both hypothetical datasets, to choose the superior experimental design. We then use the model structure to interpret, and subsequently resolve, poor identifiability. This concludes the conceptual aspect of the analysis. The following four paragraphs address the concrete aspects, being a) generating the two hypothetical datasets, b) the investigated parameters, c) the hypothesis, and d) the assessment criteria used to evaluate the hypothesis.

Generating two hypothetical datasets Both hypothetical datasets are generated from the same forward run of the reaction network. For this purpose, we use the forward run initialised in Section 5.1 with the current baseline values. For Experiment Design A we simulate this specific situation over thirty days. The simulation for Experimental Design B is limited to three days. The output of the forward run is then processed to mimic the measurement points of the corresponding experiments. Table 18 shows that both experimental designs would monitor the pH, the cumulative concentration of CO_2 and the concentration of cellulose, lignin, monosaccharides, monoaromatic compounds, NAM and biomass. Additionally, Experimental Design A would also monitor the concentration of peptidoglycan, whereas Experimental Design B would monitor the concentration of NAG. These variables are all direct outputs of the forward run. The processing is thus limited to selecting the output at discrete time steps, which correspond with the proposed sample times for the future experiments. The number of samples is given in Table 18, and are spread evenly over the proposed length of the experiment. The last step is adding a normally distributed noise to the datasets. The standard deviation of the noise is attributed based on the standard deviation of the corresponding measurement, see Table 18, due to (hypothetical) measurement errors.

Parameters under investigation As has been noted in Section 2.4, a large amount of uncertainty is linked to the maximum rate parameters k_{max} of the depolymerisation reaction. In fact, no experimental data is readily available for the depolymerisation of SOM at low temperatures

Table 18: Hypothetical measurements (SD and $n_{Samples}$)

Hypothetical Measurement	**SD**	$n_{samples}$
pH	0.05 [51]	30
CO2(cum)	0.006 [52]	30
Cellulose	0.005 [53]	10
Lignin	0.005 [53]	10
Monosaccharide	0.0005	10
Monoaromatic compounds	0.0005	10
NAM	0.0005	10
Biomass C_X	0.0005	10
Peptidoglycan (exclusively for Exp. A)	0.0005	10
NAG (exclusively for Exp. B)	0.0005	10

and without interference of stabilisation mechanisms. Consequently, we include the k_{max} for the depolymerisation reaction of all polymers in the reaction network. Furthermore, $k_{max,decay}$ of biomass decay is incorporated in the analysis as well, because it can adopt a large range of values. The last factor of respectable uncertainty is the biomass growth on LCFA and monoaromatic compounds. This uncertainty arises from the fact that LCFA and monoaromatic compounds are atypical monomers. With this in mind, we include μ_{max} and $Y_{X/S}$ of biomass growth on LCFA and monoaromatic compounds, as well as $K_{S,Ph}$ and $K_{i,Ph}$, the inhibition parameters of phenol substrates. The chosen parameters are summarised rows 1 to 15 of Table 13. The value ranges within which DREAM searches for an optimal value are the ones suggested in Table 13 as well. DREAM is then run for 200.000 iterations for each of the datasets, optimising the above-mentioned parameters within the corresponding ranges.

Hypothesis It is hypothesised that, due to the difference in simulation time, the parameters related to the MOM fraction (e.g. depolymerisation of chitin) will be better identifiable in the dataset based on Experimental Design A. As an additional measurement, experiment A monitors the concentration of peptidoglycan, whereas Experiment B monitors the concentration of NAG. As a result, no microbial polymer is measured in experiment B. We hypothesise that there will be a larger correlation between the decay rates and the depolymerisation rates of microbial residues in experimental design B. In general, we expect that it will be possible to interpret the correlations using the inherent model structure.

Assessment criteria To be able to assess the identifiability of the parameters we will generate and discuss several results:

1. The hypothetical measurement points versus the simulated data.
2. The joint posterior distributions. The dataset has a larger sensitivity to the parameters with narrower joint posterior distributions $P(\theta|\hat{y})$.
3. The Kullback-Leibler Divergence (D_{kl}) criterion uses the joint prior and posterior distribution to quantify and compare the identifiability of the parameters from both experimental

set-ups. D_{kl} is defined with Eq. 10:

$$D_{kl} = \int (p(\theta_i|\hat{y}) \cdot ln\left(\frac{p(\theta_i|\hat{y})}{p(\theta_i)}\right) \cdot d\theta_i \tag{10}$$

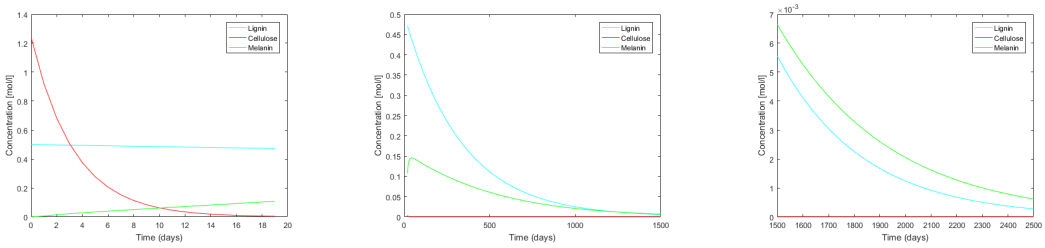
4. The correlation plots. When parameters are strongly correlated, it is not possible to assess the sensitivity of the dataset to the parameters separately.

6 Results

This section contains the results of the methods described in Section 5. The discussion of the results is done in Section 7.

6.1 Simulations by forward run with baseline values

In this section we present the results of the 2500 days simulation with baseline values. Figure 5 displays the course of the cellulose, lignin and melanin concentration. The long simulation period has been divided in three stage, as follows: a) from day 0 to day 20, b) from day 20 to day 1500, and c) from day 1500 to day 2500 (Figures 5a, 5b and 5c). The division in stages facilitates the comparison of the compound concentration in relative terms. In Figure 5a (Stage I) the concentration of cellulose starts higher than that of lignin, but rapidly decreases. After 4 days the cellulose concentration becomes less than that of lignin. At the start of the simulation, there is no melanin present in the system. It increases slowly throughout stage I. In Figure 5b (Stage II), it can be seen that cellulose has vanished from the system (or SOM mixture). Lignin and melanin are now following a similar course of concentration decrease, and approaching each others value. Figure 5c shows that melanin and lignin have a similar concentration, where lignin is slightly higher.

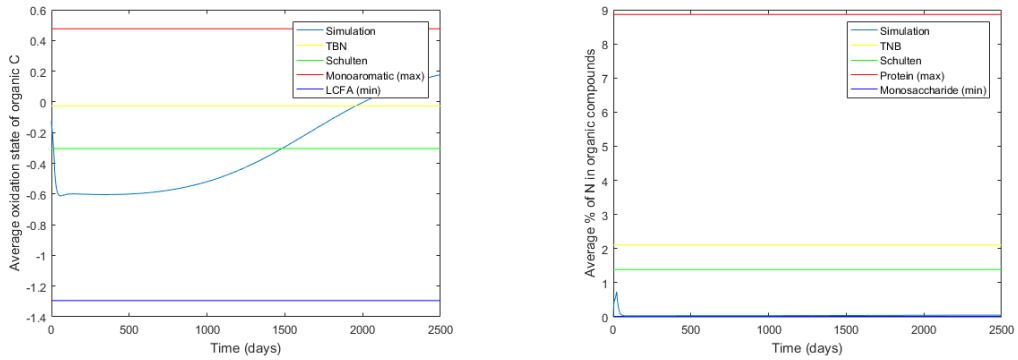


(a) From day 0 to day 20 **(b)** Fom day 20 to day 1500 **(c)** From day 1500 to day 2500

Figure 5: Concentration of cellulose, lignin and melanin from day 0 to day 2500

The figures above are direct results of the concentration of the individual compounds. In addition, we can present results of bulk properties of SOM by combining those concentrations with the properties of the individual compounds. Figure 6 presents the average oxidation state of the carbon atoms in the SOM mixture and the elemental analysis for nitrogen. Both properties can be directly deducted from the chemical formula of the individual compounds and their concentration. In Figure 7, we show the average amount of carboxyl and phenolic groups per carbon atom in the mixture. In order to calculate these two properties, we need the chemical formula, the amount of carboxyl and phenolic groups and the concentration for each individual compound. In the following four paragraphs, we will describe the simulation of each simulation in more detail.

Figure 6a The average oxidation state of the carbon atoms in any SOM mixture in our model always lies between -1.29 and 0.48. These limits correspond with the oxidation state of C in resp. LCFA and monoaromatic compounds. The actual range of values covered by the SOM mixture



(a) Average oxidation state of the organic carbon atoms

(b) Average percentage of N atoms in an organic compound

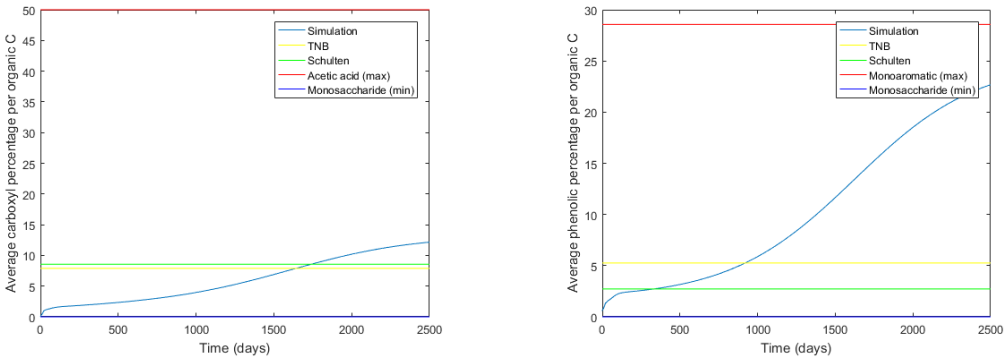
Figure 6: Average property of bulk SOM

is [-0.6;0.2]. The average oxidation state of carbon in TNB and the Schulten SOM molecule fall within the simulated range of values. A SOM mixture which has the same oxidation state as the TNB molecule is achieved after approximately 1900 days. For the Schulten molecule this is the case after approximately 1500 simulated days.

Figure 6b The average percentage of N in any SOM mixture in our model always lies between 0% and 8.9%. This corresponds with the composition of resp. monosaccharide and protein. The actual range of values covered by the SOM mixture is much narrower, i.e. [0%;0.8%]. The average percentages of N in TNB (2.1%)and the Schulten SOM molecule (1.39%) do not fall within the simulated range of values.

Figure 7a The average amount of carboxyl groups per carbon atoms in any SOM mixture in our model always lies between 0% and 50%. These limits correspond with the amount of carboxyl groups in C in resp. monosaccharides and acetic acid. The actual range of values covered by the SOM mixture is [0%;12%]. The average amounts of carboxyl groups per C atoms in TNB (7.89%) and the Schulten SOM molecule (8.56%) fall within the simulated range of values. A SOM mixture which has the same average amount of carboxyl group as the TNB molecule is reached after approximately 1650 days. For the Schulten molecule this is the case after approximately 1750 simulated days.

Figure 7b The average amount of phenolic groups per carbon atoms in any SOM mixture in our model always lies between 0% and 28.5%. These limits correspond with the amount of carboxyl groups in C in resp. monosaccharides and monoaromatic compounds. The actual range of values covered by the SOM mixture is [1%;22.5%]. The average amounts of phenolic groups per C atoms in TNB (5.26%) and the Schulten SOM molecule (2.72%) fall within the simulated range of values. A SOM mixture which has the same average amount of phenolic group as the TNB molecule is reached after approximately 900 days. For the Schulten molecule this is the case after approximately 300 simulated days.



(a) Average percentage of carboxyl groups per organic C

(b) Average percentage of phenolic groups per organic C

Figure 7: Average property of bulk SOM

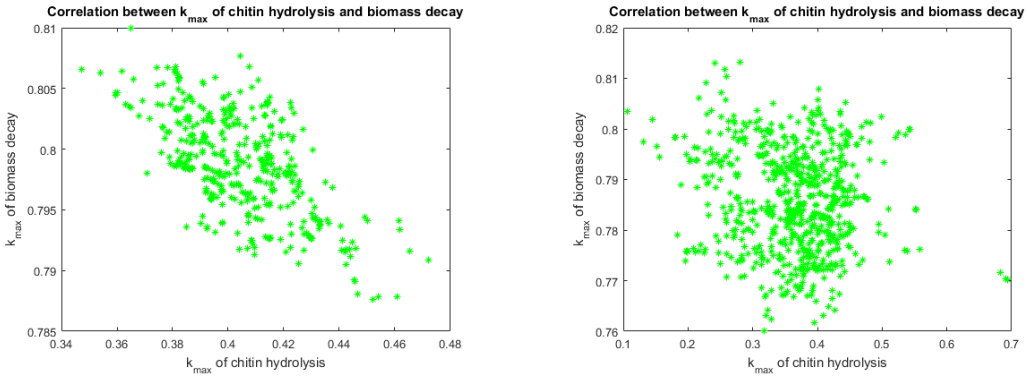
6.2 Results of experimental design selection

Identifiability (D_{kl}) and posterior ranges The identifiability, quantified by D_{kl}, and posterior ranges of the parameters which have converged in at least one experimental design, are given in Table 19. The experimental design with the largest D_{kl} is indicated in bold for each value. The D_{kl} of $k_{max,cellulose}$ is the highest in experimental design B. For all other parameters, the D_{kl} is the highest in experimental design A. Some posterior ranges do not include the initial hypothetical value.

Table 19: Posterior ranges (5% - 95%) and D_{kl}

	$k_{cellulose}$ k_{hyp} = 0.3		k_{lignin} k_{hyp}=0.003		$k_{protein}$ k_{hyp} = 0.3		k_{chitin} k_{hyp} = 0.3	
Set-up	Quantile	D_{kl}	Quantile	D_{kl}	Quantile	D_{kl}	Quantile	D_{kl}
A	[0.2963;0.3053]	5.63	[0.0013;0.0021]	**3.28**	[0.2556;0.2969]	**3.92**	[0.3601;0.4330]	**3.49**
B	[0.2989;0.3036]	**6.26**	[0.0011;0.0024]	2.86	[0.6585;1.0313]	1.80	[0.2380;0.4582]	2.26
	k_{glucan} k_{hyp}=0.3		$k_{peptidoglycan}$ k_{hyp}=0.003		$k_{bacteran}$ k_{hyp}=0.003		$k_{melanin}$ k_{hyp}=0.003	
Set-up	Quantile	D_{kl}	Quantile	D_{kl}	Quantile	D_{kl}	Quantile	D_{kl}
A	[0.2895;0.3120]	**4.62**	[0.2624;0.3029]	**4.00**	[0.0063;0.0144]	1.04	[0.0009;0.0035]	**2.31**
B	[0.2983;0.4413]	2.70	[0.0487;0.1686]	3.15	[0.0088;0.0287]	0.32	[0.0176;0.0297]	0.66
	$\mu_{monoaromatic}$ k_{hyp}=0.3		k_{decay} k_{hyp}=0.8		$K_{s,monoaromatic}$ k_{hyp} = 0.01			
Set-up	Quantile	D_{kl}	Quantile	D_{kl}	Quantile	D_{kl}		
A	[1.5965;2.9816]	**0.87**	[0.7927;0.8057]	**7.06**	[0.2434;0.4894]	**1.16**		
B	[2.5722;2.9674]	1.86	[0.7684;0.8029]	6.02	[0.2416;0.7455]	0.49		

Correlations Lastly, we analysed correlations between parameters for both reaction networks. Figures 8a and 8b show the correlation between the k_{max} of chitin hydrolysis and that of biomass decay for both experimental designs. We observed a strong correlation between the two parameters in experimental design A, but not in experimental design B.

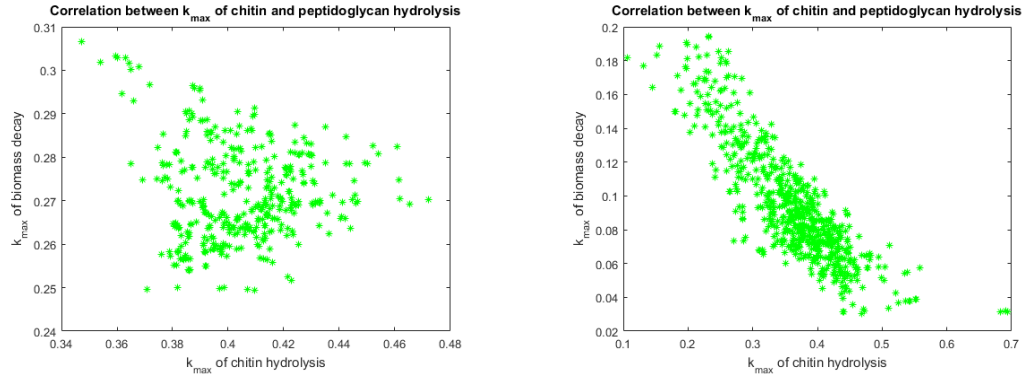


(a) Correlation between k_{max} of chitin and biomass decay for Experimental Design A

(b) Correlation between k_{max} of chitin and biomass decay for Experimental Design B

Figure 8: Correlation between k_{max} of chitin and biomass decay after 200.000 iteration runs

Figures 9a and 9a represent the correlation between k_{max} of chitin and peptidoglycan hydrolysis. In experimental design A, a correlation is absent. On the other, we observed a correlation between the two parameters for experimental design B.

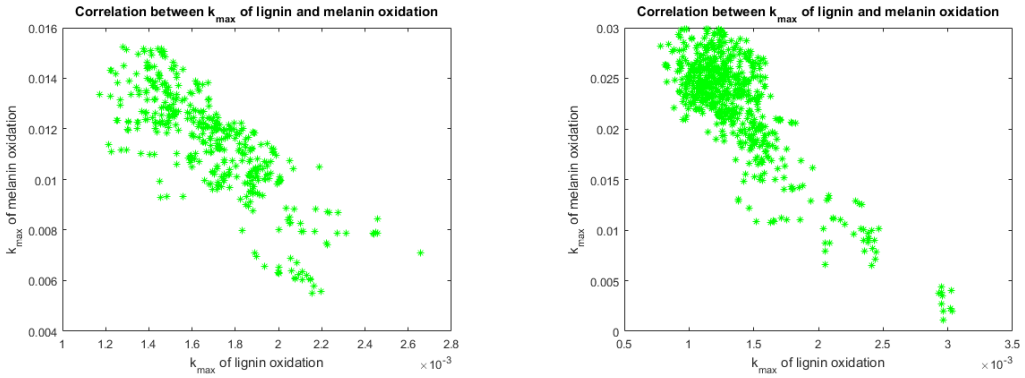


(a) correlation between k_{max} of chitin and peptidoglycan hydrolysis for Experiment A

(b) Correlation between k_{max} of chitin and peptidoglycan hydrolysis for Experiment B

Figure 9: Correlation between k_{max} of chitin and peptidoglycan hydrolysis after 200.000 iteration runs

The correlations between k_{max} of lignin and melanin oxidative depolymerisation is displayed in Figures 10a and 10b. For both experimental designs, correlation can be observed.



(a) Correlation of k_{max} of lignin and melanin for Experiment A

(b) Correlation of k_{max} of lignin and melanin for Experiment B

Figure 10: Correlation of k_{max} of lignin and melanin after 200.000 iteration runs

7 Discussion

This section discusses the results presented in Section 6.

7.1 Discussion of simulations with baseline values

The three stages show in Figure 5 can roughly be interpreted as follows:

- **Stage I:** A stage where the SOM mixture is governed by plant residues, with both easily degradable (e.g. cellulose) and relatively resistant (e.g. lignin) plant compounds. This stage ends rapidly (after 4 days) when the easily degradable fraction has been degraded (Figure 5a).
- **Stage II:** A stage where the SOM mixture is governed by relatively resistant plant compounds (e.g. lignin), and a growing fraction of resistant microbial residues (e.g. melanin). This stage spans from day 4 to day 1500, and is thus more than 350 times longer than stage I (Figure 5b).
- **Stage III:** A stage where the SOM mixture is governed by resistant plant compounds (e.g. lignin) and resistant microbial residues (e.g. melanin). This stage spans at least until day 2500 and is expected to be maintained even longer (Figure 5c).

For the properties of bulk SOM, the absolute concentration of individual compounds is not important. It is the relative concentration that governs the bulk properties.

For all studied bulk properties, i.e. average oxidation state, % of N, % of carboxyl groups and % of phenolic groups, the model can technically simulate SOM mixture that have the corresponding value of TNB and Schulten SOM molecule (Figures 6 and 7). For all properties, except the percentage of N, this SOM mixture can found with the mixture simulated with the current initialisation and simulation time (Figures 6a, 7a and 7b.

The average percentage of N of the SOM mixture stays lower than that of TNB and Schulten SOM during the whole simulation (Figure 6b. However, the values do fall within the technical limitations of the property. This observation could thus be explained by factors that do not question the validity of the compounds and reaction types in the reaction network itself. A SOM mixture that does have the same property could be reached by adapting the initial conditions, the redistribution of necromass into several microbial polymers or one or more reaction rates. Another possibility is that we are either missing a nitrogen rich compound in our mixture, or that the amount of N of one or more compounds has been underestimated.

For the oxidation state of C and for the amount of carboxyl groups, the simulation time at which the correct values are met lies within Stage III. These SOM mixture thus contain mostly microbial residues and resistant plant residues. On the other hand, for the phenolic groups to match the phenolic groups in TNB and Schulten, we need a SOM mixture from Stage I. This SOM mixture thus correspond to a mixture of both easily degradable and resistant plant compounds. This is not necessarily contradictory, because the model humic molecules are themselves representations of several SOM mixtures, from which experimental data has been published. An interesting future analysis would be to compare the simulations to the experimental data of one specific SOM mixture. In that case, the same simulated SOM mixture should give reasonable results for the several properties.

7.2 Discussion of experimental design selection

In this section, we will discuss the suitability of the two experimental designs A and B, based on the identifiability of parameters based on D_{kl} and correlations, shown in Section 6.2.

From Table 19 it can be seen that set-up A is better at identifying (i.e. higher D_{kl}) the depolymerisation rate of protein, chitin and peptidoglycan as well as the decay rate and the growth yield on monoaromatics. Similarly, set-up B is better at identifying the rates related to the plant fraction.

The results of the analysis are only partially in accordance with the hypothesised results. The discordances highlight the importance of running an automatised analysis of the experimental designs and use clearly defined statistical parameters to compare the designs. The best experimental design might not be the one which would be chosen intuitively. Especially for complex reaction networks with high numbers of fitted parameters, it is important to do so. Furthermore, the analysis can also reveal which adaptation can be made to the currently best experimental design, to increase the identifiability of the fitted parameters.

We hypothesised that a stronger correlation between k_{max} of microbial polymers and biomass decay would be observed for experimental design B, than for A, due to the lack of measurements monitoring microbial polymers. However, Figure 8b does not show any correlation, while 8a does. A possible explanation is that the microbial fraction is already poorly identifiable with Experimental Design B, which would influence the identifiability of the correlations.

In 9b a correlation between k_{max} of chitin and peptidoglycan hydrolysis is noted. This correlations finds its origin in the similar composition of chitin and peptidoglycan. Both polymers release NAM. Due to the additional measurement of peptidoglycan in Experimental Design A, the correlation disappears.

The strong correlation between the k_{max} of lignin and melanin was to expected, because both polymers have been characterised in the same way: same chemical formula, same properties and same k_{max}. Therefore, the correlation can only be removed by monitoring directly one of the two concentrations.

8 Conclusions

The main context from which this study evolved, is the urgent need for a versatile mechanistic modelling tool that can address the ever-growing diversity in research questions related to SOM degradation. The chosen starting point for the development of such a tool, was to look for the common denominator across all the research questions. The result of this paper is a mechanistic toolbox, which features both a forward model for SOM degradation as well as a tool for Bayesian inference analysis, i.e. DREAM. The forward model is the implementation of an aerobic SOM degradation network, where SOM is defined as a mixture of known organic compounds of plant and microbial origin, which interacts with microorganisms in a non-limiting aqueous environment.

The major conclusions that can be drawn based on the development and evaluation of our novel mechanistic toolbox, are presented below:

1. The current literature suffices to make informed and justified choices with regard to defining the main SOM degradation compounds and reactions, along with their stoichiometry. However, the literature lacks the data to adequately quantify the kinetics of such degradation reactions. The reason for this can be found in the novel nature of our approach: so far, the kinetic parameters of certain reactions have rarely been studied under non-limiting conditions combined low pressures and temperatures.
2. During the numerical implementation of the reaction network it became apparent that we can deal with the uncertain nature of the kinetic parameters, by incorporating them as input variables, rather than hard-coding their value in the main body of the forward model. As such, the standard values, i.e. baseline values, can be easily updated whenever new information is obtained. Furthermore, the uncertain parameters can then be analysed with DREAM.
3. To implement the reaction network, the only strictly necessary characterization of the individual compounds is that of their chemical composition through a chemical formula. Other useful properties are $\mathbf{\Delta G}_f^\circ$ and the average oxidation state of the carbon atoms. The two latter properties can be used to estimate growth yields and rates. Moreover, the model structure allows us to link any non-essential property as we wish.
4. By adding noise to simulations of the forward model, we can generate hypothetical datasets that mimic real datasets with only measurement errors. A hypothetical dataset can then be combined with DREAM, or Bayesian inference in general, to gain a powerful tool that assesses the sensitivity of certain output parameters to input variables, as well as correlations, which can indeed be related back to the inherent model structure.
5. The combination of a hypothetical dataset and the DREAM module can be deployed to select experimental designs which take into account avoidable poor identification of the parameters of interest, by assess the model structure for sensitivity and correlations.

With these insights, the toolbox can already be a valuable asset in further research. We describe two applications that can be initiated in a near future with the toolbox as such:

1. The most immediate application of the toolbox is to define the first experiment that should be conducted to gather data which can identify the uncertain parameters. The discussion

in Section 7.2 illustrates how this should be done. This process needs to be iterated several times. Once the first experiment has been conducted, the DREAM needs to be used to assess the real identifiability of the parameters. At this point, a new iteration process will start, since with a real dataset there is the additional influence of model errors.

2. The second proposed application is to use several mixtures of SOM, defined by individual compounds, as input for molecular simulations [48, 54]. Currently, those molecular simulations are mostly applied on representative DOM molecules, such as the Schulten Molecule [48]. These molecules are characterised with properties that reflect experimentally determined properties of humic substances. Due to the flexible nature of our toolbox, similar properties can be added to the individual compounds. During the simulations we could then study the supramolecular association behaviour of the compounds. This can help reinterpret older experimental data, in terms of humic substances, in the light of the most recent view on SOM and its degradation.

To conclude this report, we address a (non-exhaustive) list of recommended improvements on the toolbox itself:

1. The most profound adaptation to the current reaction network, will be addition of anaerobic SOM degradation. This will result in new SOM compounds, especially with regard to the byproducts originating from substrate growth [35,55–57]. The addition of an anaerobic network will require an extended literature study, but its implementation will be straightforward due to the model flexibility. The true difficulty might be to cope with the increasing depletion of oxygen, which will not create a distinct switch between aerobic and anaerobic conditions. A starting point for an anaerobic reaction network can be the Anaerobic Digestion Model [35] for wastewater treatment.

2. All biomass types are currently lumped together to form one component. It is recommended to add more types of biomass, starting with a distinction between fungal and bacterial biomass. The proportion of fungal and bacterial biomass is important for the resulting necromass composition. At the moment, this proportion is set constant throughout time. Another reason for adding multiple biomass types, is that not all polymers and monomers can be degraded by all biomass. When for example no white-rot fungi is present in the system, the lignin decomposition will come to a halt. In anaerobic conditions, where only specialised biomass is capable of degrading compounds with high recalcitrance, this concept becomes increasingly important.

3. The general process of microbial growth could also be expanded to better mimic the concept of microbial growth efficiency (MGE). Modelling extracellular polymeric substances can be used to model biomfilms, which in turn influence the uptake rate of monomers [31]. Furthermore, the principle of MGE is linked to the availability of nitrogen [31]. In the current version, depletion of NH_3 or N-containing substrates decreases the uptake rate of substrates. However, a more realistic process would be to have equivalent uptake rates, but a decreased MGE, with higher production of CO_2 and other byproducts [31].

How large is the contribution of microbial organic matter to the total organic matter? Which type of organic matter mixture will result in maximum mineral binding? What is the temperature dependence of stable OM pools [58]? With our novel toolbox, we are now one step closer to numerically approach this wide range of hypotheses.

References

[1] M. W. I. Schmidt, M. S. Torn, S. Abiven, T. Dittmar, G. Guggenberger, I. A. Janssens, M. Kleber, I. Kögel-Knabner, J. Lehmann, D. A. C. Manning, P. Nannipieri, D. P. Rasse, S. Weiner, and S. E. Trumbore, "Persistence of soil organic matter as an ecosystem property," *Nature*, vol. 478, no. 7367, pp. 49–56, 2011. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature10386

[2] J. Lehmann, D. Solomon, J. Kinyangi, L. Dathe, S. Wirick, and C. Jacobsen, "Spatial complexity of soil organic matter forms at nanometre scales," *Nature Geoscience*, vol. 1, no. 4, pp. 238–242, 2008. [Online]. Available: http://www.nature.com/doifinder/10.1038/ngeo155

[3] M. Segoli, S. De Gryze, F. Dou, J. Lee, W. Post, K. Denef, and J. Six, "Soil Aggregate Dynamics : A Mechanistic Model," *Ecological Modelling*, vol. 263, pp. 1–9, 2011.

[4] K. Coleman and D. Jenkinson, "RothC - A Model for the Turnover of Carbon in Soil," *Evaluation of Soil Organic Matter Models: Using Existing Long-Term Datasets*, no. June, pp. 237–246, 1996.

[5] S. Manzoni and A. Porporato, "Soil carbon and nitrogen mineralization: Theory and models across scales," *Soil Biology and Biochemistry*, vol. 41, no. 7, pp. 1355–1379, 2009. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038071709000765http://dx.doi.org/10.1016/j.soilbio.2009.02.031

[6] T. J. Heimovaara, "A toolbox to find the best mechanistic model to predict the behavior of environmental systems ." vol. 83, pp. 344–355, 2016.

[7] M. v. Lutzow, I. Kogel-Knabner, K. Ekschmitt, E. Matzner, G. Guggenberger, B. Marschner, and H. Flessa, "Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions - a review," *European Journal of Soil Science*, vol. 57, no. 4, pp. 426–445, 2006. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2389.2006.00809.x

[8] M. E. Essington, *Soil and Water Chemistry - an integrative approach*, 2004, no. 1.

[9] E. Tipping, *Cation Binding by Humic Substances*. Cambridge: Cambridge University Press, 2002. [Online]. Available: http://arxiv.org/abs/1011.1669http://ebooks.cambridge.org/ref/id/CBO9780511535598

[10] B. P. Kelleher and A. J. Simpson, "Humic substances in soils: Are they really chemically distinct?" *Environmental Science and Technology*, vol. 40, no. 15, pp. 4605–4611, 2006.

[11] M. KLEBER, P. S. NICO, A. PLANTE, T. FILLEY, M. KRAMER, C. SWANSTON, and P. SOLLINS, "Old and stable soil organic matter is not necessarily chemically recalcitrant: implications for modeling concepts and temperature sensitivity," *Global Change Biology*, vol. 17, no. 2, pp. 1097–1107, 2011. [Online]. Available: http://doi.wiley.com/10.1111/j.1365-2486.2010.02278.x

[12] W. J. Riley, F. Maggi, M. Kleber, M. S. Torn, J. Y. Tang, D. Dwivedi, N. Guerry, and C. W. J. Riley, "Long residence times of rapidly decomposable soil organic matter : application of a multi-phase , multi-component , and vertically resolved model (BAMS1) to soil carbon dynamics," pp. 1335–1355, 2014.

[13] J. Six, S. D. Frey, R. K. Thiet, and K. M. Batten, "Bacterial and Fungal Contributions to Carbon Sequestration in Agroecosystems," *Soil Science Society of America Journal*, vol. 70, no. 2, p. 555, 2006.

[14] I. Kögel-Knabner, "The macromolecular organic composition of Plant and microbial residues as inputs to soil organic matter," *Soil Biology and Biochemistry*, vol. 34, no. 2, pp. 139–162, 2002.

[15] S. E. Jacobsen and C. E. Wyman, "Cellulose and hemicellulose hydrolysis models for application to current and novel pretreatment processes." *Applied biochemistry and biotechnology*, vol. 84-86, pp. 81–96, 2000.

[16] M. Busse-Wicher, T. C. F. Gomes, T. Tryfona, N. Nikolovski, K. Stott, N. J. Grantham, D. N. Bolam, M. S. Skaf, and P. Dupree, "The pattern of xylan acetylation suggests xylan may interact with cellulose microfibrils as a twofold helical screw in the secondary plant cell wall of Arabidopsis thaliana," *Plant Journal*, vol. 79, no. 3, pp. 492–506, 2014.

[17] T. D. H. Bugg, M. Ahmad, E. M. Hardiman, and R. Rahmanpour, "Pathways for degradation of lignin in bacteria and fungi." *Natural Product Reports*, vol. 28, no. 12, pp. 1883–96, 2011. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/21918777$\backslash$nhttp://pubs.rsc.org/en/content/articlehtml/2011/np/c1np00042j

[18] T. K. Bhat, B. Singh, and O. P. Sharma, "Microbial degradation of tannins – A current perspective - Springer," no. Graham 1992, pp. 343–357, 1998.

[19] Q. Huang, P. M. Huang, and A. Violante, *Soil Mineral–Microbe-Organic Interactions*, 2008. [Online]. Available: http://www.springer.com/geosciences/book/978-3-540-77685-7?cm{_}mmc=NBA-{_}-May-08{_}WEST{_}1843704-{_}-product-{_}-978-3-540-77685-7

[20] P. E. Kolattukudy, "Structure, Biosynthesis, and Biodegradation of Cutin and Suberin," *Annual Review of Plant Physiology and Plant Molecular Biology*, vol. 32, pp. 539–567, 1981. [Online]. Available: ⟨GotoISI⟩://WOS:A1981LT25600020$\backslash$nhttp://www.annualreviews.org/doi/pdf/10.1146/annurev.pp.32.060181.002543

[21] H. Torabizadeh, "All Proteins Have a Basic Molecular Formula," *International Scholarly and Scientific Research & Innovation*, vol. 5, no. 6, pp. 777–781, 2011. [Online]. Available: http://www.waset.org/publications/10042

[22] H. Chen, *Biotechnology of Lignocellulose*, 2014. [Online]. Available: http://link.springer.com/10.1007/978-94-007-6898-7

[23] J. Pérez, J. Muñoz-Dorado, T. de la Rubia, and J. Martínez, "Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview," *International Microbiology*, vol. 5, no. 2, pp. 53–63, 2002. [Online]. Available: http://link.springer.com/10.1007/s10123-002-0062-3

[24] C. Crestini, M. Crucianelli, M. Orlandi, and R. Saladino, "Oxidative strategies in lignin chemistry: A new environmental friendly approach for the functionalisation of lignin and lignocellulosic fibers," *Catalysis Today*, vol. 156, no. 1-2, pp. 8–22, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.cattod.2010.03.057

[25] R. Kleerebezem and M. C. M. van Loosdrecht, "A Generalized Method for Thermodynamic State Analysis of Environmental Systems," *Critical Reviews in Environmental Science and Technology*, vol. 40, no. 1, pp. 1–54, 2010.

[26] M. Farrell, P. W. Hill, J. Farrar, T. H. Deluca, K. Kielland, R. Dahlgren, D. V. Murphy, P. J. Hobbs, R. D. Bardgett, and D. L. Jones, "Oligopeptides Represent a Preferred Source of Organic N Uptake : A Global Phenomenon ?" no. 3, pp. 133–145, 2013.

[27] H. M. Throckmorton, J. A. Bird, L. Dane, M. K. Firestone, and W. R. Horwath, "The source of microbial C has little impact on soil organic matter stabilisation in forest ecosystems," *Ecology Letters*, vol. 15, no. 11, pp. 1257–1265, 2012.

[28] M. Henze, "Modeling of Aerobic Wastewater Treatment Processes," *Environmental Biotechnology: Concepts and Applications*, no. February, pp. 121–134, 2005.

[29] Q. Qing, B. Yang, and C. E. Wyman, "Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes," *Bioresource Technology*, vol. 101, no. 24, pp. 9624–9630, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.biortech.2010.06.137

[30] A. G. van Turnhout, R. Kleerebezem, and T. J. Heimovaara, "How to find the optimal mechanistic reaction network describing your data?" *In review*, 2015.

[31] M. Redmile-Gordon, R. Evershed, P. Hirsch, R. White, and K. Goulding, "Soil organic matter and the extracellular microbial matrix show contrasting responses to C and N availability," *Soil Biology and Biochemistry*, vol. 88, pp. 257–267, 2015. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0038071715001959

[32] S. E. Agarry, B. O. Solomon, and S. K. Layokun, "Substrate inhibition kinetics of phenol degradation by binary mixed culture of Pseudomonas aeruginosa and Pseudomonas fluorescence from steady state and wash- out data," *African Journal of Biotechnology*, vol. 7, no. 21, pp. 3927–3933, 2008. [Online]. Available: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-56249113368{\&}partnerID=40

[33] G. C. Okpokwasili and C. O. Nweke, "Microbial growth and substrate utilization kinetics," *African Journal of Biotechnology*, vol. 5, no. 4, pp. 305–317, 2005.

[34] G. Guggenberger, "Humification and mineralization in soils," *Microorganisms in Soils: Roles in Genesis and Functions*, vol. 3, pp. 85–106, 2005. [Online]. Available: http://link.springer.com/chapter/10.1007/3-540-26609-7{_}4http://link.springer.com/chapter/10.1007/3-540-26609-7{\{}{\{}{\}}{\{}{_}{\}}{\{}{\}}{\}}4http://link.springer.com/chapter/10.1007/3-540-26609-7{\{}{_}{\}}4

[35] D. J. Batstone, "Anaerobic Digestion Model 1," vol. 52, no. S1, p. 1, 2013. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/24156076

[36] J. A. Vrugt, "Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation," *Environmental Modeling & Software*, vol. 75, pp. 273–316, 2016. [Online]. Available: http://dx.doi.org/10.1016/j.envsoft.2015.08.013

[37] J. J. Heijnen, "Impact of Thermodynamic Principles in Systems Biology." Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 139–163. [Online]. Available: http://link.springer.com/10.1007/10{_}2009{_}63

[38] PubChem, "palmitic acid — C16H32O2 - PubChem." [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/compound/palmitic{_}acid{\#}section=Surface-Tension

[39] R. K. Thauer, K. Jungermann, and K. Decker, "Energy conservation in chemotrophic anaerobic bacteria." *Bacteriological reviews*, vol. 41, no. 1, pp. 100–180, 1977.

[40] "Escherichia coli K-12 substr. MG1655 protocatechuate." [Online]. Available: http://www.biocyc.org/ECOLI/NEW-IMAGE?type=NIL{\&}object=3-4-DIHYDROXYBENZOATE{\&}redirect=T

[41] "16-Hydroxyhexadecanoic acid (CAS 506-13-8) - Chemical & Physical Properties by Cheméo." [Online]. Available: https://www.chemeo.com/cid/82-277-4/16-Hydroxyhexadecanoicacid

[42] "Escherichia coli K-12 substr. MG1655 N-acetyl-D-glucosamine." [Online]. Available: http://www.biocyc.org/ECOLI/NEW-IMAGE?type=COMPOUND{\&}object=N-acetyl-D-glucosamine

[43] "Escherichia coli K-12 substr. MG1655 N-acetylmuramate." [Online]. Available: http://biocyc.org/ECOLI/NEW-IMAGE?type=COMPOUND{\&}object=NACMUR

[44] M. Schweigert, S. Herrmann, A. Miltner, T. Fester, and M. Kästner, "Fate of ectomycorrhizal fungal biomass in a soil bioreactor system and its contribution to soil organic matter formation," *Soil Biology and Biochemistry*, vol. 88, pp. 120–127, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038071715001820

[45] G. Koch, M. Kühni, W. Gujer, and H. Siegrist, "Calibration and validation of activated sludge model no. 3 for Swiss municipal wastewater," *Water Research*, vol. 34, no. 14, pp. 3580–3590, 2000.

[46] R. Gupta, S. Kumar, J. Gomes, and R. C. Kuhad, "Kinetic study of batch and fed-batch enzymatic saccharification of pretreated substrate and subsequent fermentation to ethanol," *Biotechnology for Biofuels*, vol. 5, no. 1, p. 16, 2012. [Online]. Available: http://www.biotechnologyforbiofuels.com/content/5/1/16

[47] M. F. Cotrufo, J. L. Soong, A. J. Horton, E. E. Campbell, M. L. Haddix, D. H. Wall, and W. J. Parton, "Formation of soil organic matter via biochemical and physical pathways of litter mass loss," *Nature Geoscience*, vol. 8, no. 10, pp. 776–779, sep 2015. [Online]. Available: http://www.nature.com/doifinder/10.1038/ngeo2520

[48] R. Sutton, G. Sposito, M. Diallo, and H.-R. Schulten, "Molecular simulation of a model of dissolved organic matter," *Environmental Toxicology and Chemistry*, vol. 24, no. 8, pp. 1902–1911, 2005.

[49] M. Orsi, "Molecular dynamics simulation of humic substances," *Chemical and Biological Technologies in Agriculture*, vol. 1, no. 1, p. 10, 2014. [Online]. Available: http://www.chembioagro.com/content/1/1/10

[50] H. Schulten and M. Schnitzer, "A State of the Art Structural Concept for Humic Substances Identification of the Sex Pheromone of an Ant ,," *Naturwissenschaften*, vol. 30, pp. 29–30, 1993.

[51] E. Kardash, I. Kuselman, I. Pankratov, and S. Elhanan, "Proficiency testing of pH and electrolytic conductivity measurements of groundwater: a case study of the difference between consensus and metrologically traceable values," *Accreditation and Quality Assurance*, vol. 18, no. 5, pp. 373–381, oct 2013. [Online]. Available: http://link.springer.com/10.1007/s00769-013-1006-7

[52] A. Drolc, P. Djinović, and A. Pintar, "Gas chromatography analysis: Method validation and measurement uncertainty evaluation for volume fraction measurements of gases in simulated reformate gas stream," *Accreditation and Quality Assurance*, vol. 18, no. 3, pp. 225–233, 2013.

[53] D. W. Templeton, E. J. Wolfrum, J. H. Yen, and K. E. Sharpless, "Compositional Analysis of Biomass Reference Materials: Results from an Interlaboratory Study," *BioEnergy Research*, pp. 303–314, 2015. [Online]. Available: http://link.springer.com/10.1007/s12155-015-9675-1

[54] R. Sutton and G. Sposito, "Critical Review Molecular Structure in Soil Humic Substances : The New View," *Environmental Science and Technology*, vol. 39, no. 23, pp. 9009–9015, 2005.

[55] S. Kato, K. Chino, N. Kamimura, E. Masai, I. Yumoto, and Y. Kamagata, "Methanogenic degradation of lignin-derived monoaromatic compounds by microbial enrichments from rice paddy field soil," *Scientific Reports*, vol. 5, no. August, p. 14295, 2015. [Online]. Available: http://dx.doi.org/10.1038/srep14295$\backslash$nhttp://www.nature.com/articles/srep14295http://www.nature.com/articles/srep14295

[56] P. C. Stolk, R. F. A. Hendriks, C. M. J. Jacobs, E. J. Moors, and P. Kabat, "Modelling the effect of aggregates on N 2O emission from denitrification in an agricultural peat soil," *Biogeosciences*, vol. 8, no. 9, pp. 2649–2663, 2011. [Online]. Available: http://www.biogeosciences.net/8/2649/2011/

[57] C. Rumpel and I. Kögel-Knabner, "Deep soil organic matter-a key but poorly understood component of terrestrial C cycle," *Plant and Soil*, vol. 338, no. 1, pp. 143–158, 2011. [Online]. Available: ⟨GotoISI⟩://WOS:000285389300012$\backslash$nhttp://download.springer.com/static/pdf/55/art:10.1007/s11104-010-0391-5.pdf?auth66=1396775997{_}8170e1e5db3d5d9f6321fc903ac23ccc{\&}ext=.pdf$\backslash$nhttp://download.springer.com/static/pdf/55/art:10.1007/s11104-010-03

[58] M. von Lützow and I. Kögel-Knabner, "Temperature sensitivity of soil organic matter decomposition-what do we know?" *Biology and Fertility of Soils*, vol. 46, no. 1, pp. 1–15, 2009.