Equifinality, Sloppiness and Emergent Minimal Structures of Biogeochemical Models

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

Fakultät Naturwissenschaften
Universität Hohenheim

Institut für Bodenkunde und Standortslehre

vorgelegt von

Gianna Lisa Marschmann

aus *Oberhausen*

2018

Dekan:	Prof. Dr. rer. nat. Uwe Beifuß
1. berichtende Person:	Prof. Dr. rer. nat. Thilo Streck
2. berichtende Person:	Prof. Dr. Philipp Kügler
Eingereicht am:	05.10.2018
Mündliche Prüfung am:	18.12.2018

Summary

Process-based biogeochemical models consider increasingly the control of microorganisms on biogeochemical processes. These models are used for a number of important purposes, from small-scale (mm-cm) controls on pollutant turnover to impacts of global climate change. A major challenge is to validate mechanistic descriptions of microbial processes and predicted *emergent* system responses against experimental observations. The validity of model assumptions for microbial activity in soil is often difficult to assess due to the scarcity of experimental data. Therefore, most complex biogeochemical models suffer from *equifinality*, i.e. many different model realizations lead to the same system behavior. In order to minimize parameter *equifinality* and prediction uncertainty in biogeochemical modeling, a key question is to determine what can and cannot be inferred from available data. My thesis aimed at solving the problem of *equifinality* in biogeochemical modeling. Thereby, I opted to test a novel mathematical framework that allows to systematically tailor the complexity of biogeochemical models to the information content of available data.

The conceptual part of this thesis investigates the problem of parameter inference by nonlinear least squares regression analysis for five selected biogeochemical models of different complexity from the literature. The inverse problem proved to be ill-posed for all considered models, even if the calibration data was continuous and essentially noiseless. I found that all models show *sloppiness*, i.e. model sensitivities against parameter changes are evenly distributed over several orders of magnitude. *Sloppiness* has been identified as a characteristic feature of multi-parameter models and effective theories in systems biology and physics. Here, I discussed practical consequences of *sloppiness* for biogeochemical modeling. Based on a geometric interpretation of nonlinear least squares regression analysis, I demonstrated the impact of *sloppiness* on parameter estimation, model simplification and model predictions for a minimal biogeochemical model.

As an alternative to complete model calibration, short-term substrate-induced heterotrophic respiration analysis is widely used to independently estimate selected biophysiological parameters of biogeochemical models. I analyzed a typical model of microbial activity in soil and found that the exponentially-increasing phase of heterotrophic respiration can only be used to estimate the maximum specific growth rate and an effective parameter which was defined by a non-linear combination of the initial total and active microbial fraction as well as the true growth yield. The non-identifiability of parameters could be

attributed to *sloppiness*. I showed that a geodesic-following algorithm that exploits the *sloppy* structure of parameter space can be systematically used to better constrain parameters in heterotrophic respiration analysis.

The most complex biogeochemical model analyzed in this thesis describes pesticide degradation coupled to carbon turnover in soil on the millimeter scale. Using methods from information geometry, I derived a less complex model formulation with effective inferable parameters that reflects essential physicochemical and microbial controls on pesticide degradation. The effective model represents a characteristic realization out of a set of *equifinal* models that show system behavior consistent with the experimental data used for calibration. The effective inferable parameters remain expressed in terms of nonlinear parameter combinations of the original *equifinal* model. I demonstrated that the information theoretic framework for model reduction can be used to reveal structural limitations of biogeochemical process formulations. Both the full and reduced models were not able to fully predict a validation data set with different boundary conditions, indicating that the original process formulation for bacterial pesticide degradation has to be revised. I showed that model complexity systematically decreases with decreasing information content in the experimental data. It is mostly sufficient to describe the system by linear effective degradation rates and equilibrium assumptions for the activity of a subset of microbial functional types and carbon pools.

In contrast to the many parameters that define the underlying biogeochemical models, the system response can often be described by a few phenomenological parameters. Instead of relying on a top-down modeling approach, I demonstrated that the complex model for pesticide degradation coupled to carbon turnover can be renormalized into a minimal model with only a few effective parameters that span the degrees of freedom of a typical system response to a chemical press disturbance. The effective parameters remain expressed in terms of their mechanistic origins and can be conceptually linked to the dynamic stability and resistance of soils.

Overall, the results of my thesis have shown that (i) *sloppiness* hampers the estimation of individual parameters of biogeochemical models, (ii) uncertainty in parameter estimates and mechanistic process descriptions can be reduced by systematic model simplifications, (iii) despite *equifinality*, it is possible to analyze the inherent small-scale complexity of microbial processes and identify important regulation mechanisms in biogeochemical systems.

Zusammenfassung

Prozessbasierte Modelle des Kohlenstoffumsatzes im Boden berücksichtigen zunehmend direkt die Dynamik von mikrobiellen Gruppen und deren Auswirkung auf biogeochemische Prozesse. Der Einsatzbereich dieser Modelle reicht von kleinskaliger Modellierung (mm-cm) von
Schadstoffumsätzen im Boden bis hin zu globalen Simulationen der Folgen des Klimawandels. Eine große Herausforderung ist es, mechanistische Beschreibungen mikrobieller Prozesse und das beobachtbare *emergente* Systemverhalten zu validieren. Besonders schwierig ist die Validierung von Modellannahmen zur Aktivität einzelner mikrobieller Gruppen im Boden, weil direkte Messungen fehlen. Die meisten komplexen biogeochemischen Modelle zeigen *Äquifinalität*, d.h. viele unterschiedliche Parameterkombinationen führen zu identischen Simulationen. Um die Parameter-*Äquifinalität* und die Vorhersageunsicherheit biogeochemischer Modelle zu minimieren, ist es wichtig, den Informationsgehalt verfügbarer Messdaten für die Modellparametrisierung zu quantifizieren. Ziel meiner Dissertation war es, das Problem der *Äquifinalität* zu lösen und einen allgemeingültigen mathematischen Formalismus zu finden, in dessen Rahmen die Komplexität biogeochemischer Modelle systematisch an den Informationsgehalt verfügbarer Daten angepasst werden kann.

Der konzeptionelle Teil dieser Dissertation behandelt das Problem der Parameterschätzung in der nichtlinearen Regressionsanalyse am Beispiel von fünf biogeochemischen Modellen unterschiedlicher Komplexität. Das inverse Problem ist für diese Modelle sogar dann extrem schlecht gestellt, wenn Daten kontinuierlich und fast ohne Rauschen gemessen werden. Alle untersuchten Modelle zeigen *Sloppiness*, d.h. sie weisen eine annähernd gleichmäßige Verteilung der Modelsensitivitäten gegenüber Parameteränderungen über mehrere Größenordnungen auf. In der Systembiologie und der Physik wurde *Sloppiness* als typisches Phänomen hoch parametrisierter Modelle und effektiver Theorien identifiziert. Hier diskutiere ich die praktischen Auswirkungen von *Sloppiness* für die Anwendung bei biogeochemischen Modellen. Anhand eines minimalen biogeochemischen Modells stelle ich die Auswirkungen von *Sloppiness* auf Parameterschätzungen, Modellvereinfachungen und -vorhersagen auf Basis einer geometrischen Interpretation der Regressionsanalyse dar.

Als Alternative zur Kalibrierung vollständiger Modelle wird die sogenannte kurzzeitige Substrat-induzierte heterotrophe Respirationsanalyse eingesetzt, um einzelne biophysio-

logische Parameter biogeochemischer Modelle zu schätzen. Die Analyse eines typischen Modells der mikrobiellen Aktivität im Boden zeigte: Messungen der Bodenatmung während der exponentiell-ansteigenden Phase nach Substrat-Zugabe erlauben ausschließlich die inverse Bestimmung der maximalen spezifischen Wachstumsrate sowie eines effektiven Parameters, der sich aus einer nichtlinearen Kombination aus anfänglicher gesamter und aktiver Biomasse und dem wahren Wachstumsertrag von Mikroorganismen ergibt. *Sloppiness* liegt auch hier der fehlenden Identifizierbarkeit einzelner Parameter zugrunde. Ich zeige, dass ein Algorithmus, der den geodätischen Weg in *sloppy* Parameterräumen ausnutzt, systematisch benutzt werden kann, um die Parameterwerte aus der Respirationsanalyse besser einzugrenzen.

Das komplexeste biogeochemische Modell, das in dieser Arbeit analysiert wird, beschreibt den an den Kohlenstoffumsatz gekoppelten Pestizidabbau auf der Millimeterskala im Boden. Mithilfe von Methoden aus der Informationsgeometrie wurden einfachere, effektive Modellformulierungen hergeleitet, die essentielle physikalisch-chemische und mikrobielle Regulationsmechanismen des Pestizidabbaus verdeutlichen. Die Parametrisierung des effektiven Modells bildet das experimentell beobachtbare Systemverhalten genauso gut ab wie komplexere Modellparametrisierungen. Die effektiven Parameter ergeben sich dabei als nichtlineare Parameterkombinationen des komplexen *äquifinalen* Ausgangsmodells und sind sicher durch Kalibrierung mit den verfügbaren Daten bestimmbar. Ich zeige, dass das neue Verfahren der Modellreduktion benutzt werden kann, um strukturelle Schwächen biogeochemischer Prozessformulierungen aufzudecken. Sowohl das vollständige als auch das reduzierte Modell waren nicht in der Lage, einen Validierungsdatensatz mit unterschiedlichen Randbedingungen vorherzusagen. Dies zeigt, dass die ursprüngliche Prozessbeschreibung des direkten Pestizidabbaus durch spezifische Bakterien überarbeitet werden muss. Die Komplexität der identifizierten effektiven Modelle war abhängig von den verfügbaren Daten. Mit abnehmendem Informationsgehalt der Daten wurden systematisch weniger komplexe effektive Modelle identifiziert. Es ist größtenteils hinreichend, das System mit linearen Abbauraten sowie Gleichgewichtsannahmen für Teilmengen der funktionellen mikrobiellen Gruppen und Kohlenstoffpools zu beschreiben.

Das beobachtbare *emergente* Systemverhalten kann oftmals mit wenigen phänomenologischen Parametern beschrieben werden. Im Gegensatz zu empirischen Top-Down Ansätzen ermöglicht mechanistische Modellreduktion als Bottom-Up Ansatz die systematische Herleitung eines minimalen Pestizidabbau-Modells mit effektiven Parametern auf Basis eines komplexen prozessbasierten Modells. Die effektiven Parameter sind Kombinationen der Originalparameter und behalten so ihre mechanistische Bedeutung. Die ermittelten effektiven Parameter entsprechen den Freiheitsgraden einer typischen Systemreaktion

auf kontinuierlichen Pestizideintrag und können konzeptionell der dynamischen Stabilität und der Resistenz von Böden zugeordnet werden.

Die Ergebnisse der Arbeit zeigen: (i) *Sloppiness* verhindert die zuverlässige Schätzung einzelner Parameter in biogeochemischen Modellen. (ii) Unsicherheiten in Parameterschätzungen und mechanistischen Prozessformulierungen können durch systematische Modellreduktion minimiert werden. (iii) Trotz *Äquifinalität* erlaubt mechanistische Modellreduktion, die systemimmanente Komplexität kleinskaliger, mikrobieller Prozesse zu analysieren und die entscheidenden Regulationsmechanismen von Stoffumsätzen im Boden zu identifizieren.

Acknowledgements

I would like to thank everyone who has offered helpful suggestions, provided feedback and made the long journey to this PhD thesis exciting and enjoyable. Thanks to the German Research Foundation (DFG) for funding within the Research Training Group "Integrated Hydrosystem Modelling" and making this PhD thesis possible.

My special thanks are due to the members of my special PhD committee. I would like to thank Prof. Dr. Thilo Streck for giving me the freedom to develop my specific research interest, providing guidance and insights in crucial moments and teaching me so much about doing science that I probably have not even fully realized yet. Prof. Dr. Philipp Kügler's critical questions and incisive suggestions were challenging and helped a great deal in keeping checks and balances on my early research efforts. Prof. Dr. Olaf A. Cirpka hosted an outstanding fall school on "Uncertainty Assessment in Hydrosystem Modelling" that sparked my interest in the research described herein.

Deep gratitude is owed to Dr. Holger Pagel for laying the foundation and contributing to this work with his expertise in modeling, his ability to shape my ideas into concise sentences, and for always encouraging discussions despite a packed schedule.

I have learned a great deal from Dr. Fereidoun Rezanezhad at the University of Waterloo, Canada. First and foremost by watching him design and carry out experiments, but also from his unwaveringly positive approach towards scientific work.

I would like to thank Irene Witte for help on daily coding problems and technical issues with the workstation as well as for echoing my occasional sighs in our shared office. Thanks to Bo Xiao for expressing his interest in my work early on and to Marvin Höge for his enthusiasm to talk about model complexity and inviting me to the LS^3 days.

A big thanks in chronological order to friends along the journey, to whom I try very hard not to compare myself to: Lisa Buchauer, Malte Probst, Thomas Hugle, Asya Sarayeva, Lisa Bentner, Jana and Lena Maier-Hein, Taylor Maavara and Adrian Mellage.

And usually last, but never least, a salute to my parents for giving me the sense of security to be able to do something in which I find meaning and whose love and support I value immensely.

Curriculum Vitae

Personal Information

Name	Gianna Lisa Marschmann
Date of Birth	09.10.1989
Place of Birth	Oberhausen

Education

2009	General qualification for university entrance (**Abitur**), Franz-Haniel Gymnasium, Duisburg (Germany)
2009 - 2013	**B.Sc.** Physics, University of Heidelberg (Germany)
2013 - 2015	**M.Sc.** Physics, University of Heidelberg (Germany), Theoretical Biophysics
2015 - 2018	**PhD student**, RTG 1826 *Integrated Hydrosystem Modelling*, Institute of Soil Science and Land Evaluation, Biogeophysics (Prof. Dr. Thilo Streck), University of Hohenheim (Germany)

Work Experience

09/2017 - 02/2018	**Student teaching assistant**, Environmental Modeling, Department of Biogeophysics, University of Hohenheim
09/2015 - 02/2016	**Student teaching assistant**, Environmental Modeling, Department of Biogeophysics, University of Hohenheim
11/2013 - 05/2015	**Student research assistant**, Department of Biomedical Optics, Max-Planck-Institute for Medical Research (Heidelberg)
02/2012 - 07/2012	**Student research assistant**, Physics of Complex Systems, Institute for Theoretical Physics, University of Heidelberg
06/2015 - 02/2016	**Voluntary work**, Coaching youth basketball, SG Heidelberg-Kirchheim e.V.

Table of Contents

List of Figures

List of Tables

Chapter 1

General Introduction

"The world in all its detail is overwhelmingly complex, yet also much simpler than it looks." (Buchanan 2015). This statement, manifested in the three related concepts of *equifinality*, *sloppiness* and *emergence*, has encouraged and enabled the modeling of complex systems across different areas of science (Savenije 2001; Machta et al. 2013; Baveye et al. 2018). The interplay of the three concepts and their practical implications for biogeochemical modeling are the central theme of this thesis.

1.1 Equifinality, Sloppiness and Emergence

Modelers of complex systems have to excel at simplifying a problem in order to make it amenable to mathematical analysis whilst preserving the inherently complicated features. If the model captures the investigated behavior of the complex system, and it is able to predict other system behavior of interest, then it can be viewed as a successful representation of reality. Many such representations will likely exist, because they *emerge* from nonlinear interactions between many constituents that comprise the complex system. Since reality is too complicated to be compressed into a model, the challenges in mechanism-oriented modeling approaches (Hunt et al. 2018) are to determine which part of a system can be safely neglected and to find a way to lump many details of the system into effective higher-order process descriptions.

From a philosophical point of view, the *equifinality* thesis rejects the notion that "science [...] is supposed to be an attempt to work towards a single correct description of reality." (Beven 2006). *Equifinality* arises when many different model realizations lead to identical system behavior (Von Bertalanffy 1969). The uniqueness of a physical reality has also been dismissed by Anderson et al. (1972) who conversely introduced the concept of *universality classes*, i.e. equivalence classes of physical systems with the same phenomenological behavior that can in fact be described by the same model. Effective models in physics, which show weak dependence of system observables on microscopic details, inherit these properties. The prime example is the diffusion equation which is valid in the solid, liquid and gas states of every material (Gauss 1877) and which projects many possible molecular configurations of particles onto an effective control parameter (the diffusion constant; Fick 1855). Both concepts justify the use of simpler theories and effective models that capture the essence of a complex system, but either average

over *equifinal* or ignore many irrelevant system details (Savenije 2001). However, while (*universal*) effective models in physics result from the application or development of a mathematical theory (e.g., continuum limits that are valid at large length and/or time scales or formal renormalization group arguments; Wilson 1971), *equifinality* is known to the biogeochemical community mostly from predictive hydrological modeling (Beven 2006). Here, it is attributed to the lack of data in the process of model calibration. Beven 2006 notes that "it should, indeed, be expected because of the overparameterisation of [models] relative to the information content of observational data available for calibration of parameter values". In biogeochemical modeling (Section 1.2) *equifinality* is generally viewed as a nuisance that hampers system understanding and leads to uncertainty in model predictions (Tang and Zhuang 2008; Manzoni et al. 2016). Nonetheless, there is a desire to apply similar theoretically well-grounded averaging techniques as in physics to biogeochemical models (Wieder et al. 2015; Getz et al. 2018). Likewise, the principle of *emergence* in complex systems refers to "a reality that is less than the sum of its parts" (Rastetter and Vallino 2015). Here, nonlinear interactions and feedbacks give rise to observed behavior that is always simpler than what it could have been, given the many constituents that comprise the system. Contrary to *equifinality*, *emergence* is viewed a desirable property as it might lead to unifying principles and effective biogeochemical laws (Baveye et al. 2018).

Sloppiness underlies both predictive models from various scientific areas and effective theories in physics (Machta et al. 2013). *Sloppiness* has a precise definition in terms of the correlation matrix of parameter uncertainties in parameter identification problems by nonlinear least squares regression (Brown and Sethna 2003). Not only is the matrix extremely ill-conditioned for *sloppy* models, but its eigenvalue spectrum also has a distinct structure. The matrix eigenvalues of *sloppy* models are locally nearly evenly spaced on a log-scale (Waterfall et al. 2006; Gutenkunst et al. 2007). This implies that only a few nonlinear parameter combinations (corresponding to large eigenvalues) can be inferred from data, with each combination being progressively more unconstrained by a roughly constant factor (Transtrum et al. 2010). What is more important, Machta et al. 2013 showed that statistical inference connects to formal scaling tools and *emergent* theories in physics. For example in the continuum limit of particle random walks the model eigenvalues are log-linearly spaced. Relevant parameters (the diffusion and drift constant of the continuum description, Risken 1996) are precisely the ones that can be inferred from the coarse-grained data. Furthermore, the *emergent* parameters are those that remain equally distinguishable at coarser scales (Raju et al. 2017). *Sloppy* parameter combinations in turn are neither constrained by data nor useful for model predictions (Mattingly et al. 2018). *Sloppiness* thus explains why it is possible to ex-

tract falsifiable predictions from complex models, despite getting many system details wrong (Gutenkunst et al. 2007). *Sloppy* models appear to be complicated, but a hidden simpler model with *emergent* model structures that describe system behavior can be systematically extracted from an underlying complex representation (Transtrum and Qiu 2014). By analyzing geometric properties of *sloppy* models (Transtrum et al. 2010, 2011), it is possible to formally connect "bottom-up descriptions of complex processes with top-down inferences drawn from data, paving the way for emergent theories in physics, biology, and beyond" (Transtrum et al. 2015). The strategy of model building to start from a complex modeling ansatz followed by successive simplifications has been applied in research areas across many disciplines (classical physics: Machta et al. 2013, nuclear physics: Nikšić and Vretenar 2016, engineering: Transtrum et al. 2018 and systems biology: Transtrum and Qiu 2016; Bohner and Venkataraman 2017; Lombardo and Rappel 2017). Derived effective parameters of reduced models have been shown to reveal physically or biologically relevant mechanistic information about the system under study. Application in systems biology has led, e.g., to the identification of important controls of adaption in allosteric macromolecules (Bohner and Venkataraman 2017), mechanisms of cardiac arrhythmias (Lombardo and Rappel 2017) or minimal topologies in biochemical enzyme networks (Transtrum and Qiu 2016). Its conceptional usefulness for biogeochemical modeling will be established below.

1.2 Biogeochemical Modeling

Microbes in soil control key functions related to carbon cycling (Paul 2014), including carbon sequestration through the decomposition of soil organic matter (Schimel and Schaeffer 2012) or physicochemical stabilization of soil organic matter (Dungait et al. 2012; Miltner et al. 2012) which are important to predict the impacts of global climate change (Todd-Brown et al. 2012, 2013; Wieder et al. 2015; Luo et al. 2017). They are involved in biodegradation and persistence of pollutants in soil (Nowak et al. 2010; Ditterich et al. 2013; Banitz et al. 2013; Pagel et al. 2016; Babey et al. 2017) and enable the soil system's ability to withstand external disturbances (Standish et al. 2014; Hawkes and Keitt 2015; Schaeffer et al. 2016). However, "Reliable predictions require more than just recognition that microbes are important for a process. Fundamental, mechanistic understanding must be established from empirical studies. The mechanisms must be unified into a theoretical framework that provides a basis for quantitative mathematical models." (Allison 2017).

Soil biogeochemical models describe the flow of nutrients, e.g., carbon or nitrogen, through a system typically composed of microbial biomass, extracellular enzymes, soluble nutrients and soil organic matter (Manzoni and Porporato 2009; Wieder et al. 2015). In

a systems approach, experimental measurements of biogeochemical processes are combined with advanced mathematical tools to yield predictive biogeochemical models. The field can often build on insights from computational systems biology where a similar framework has been established (Engl et al. 2009; Villaverde and Banga 2014). Conceptually the modeling process proceeds in five steps (Engl et al. 2009). First, data are collected from experiments or data bases (see, e.g., the references in Wieder et al. 2015, Table 3 or Google's new search engine for scientific datasets[1]). Second, the components and interactions that should be included in the mathematical representation, i.e. the model structure, are defined. Third, a statistical model (McCullagh 2002; Transtrum et al. 2014) is identified which has to be consistent with the data. This is equivalent to the estimation of parameter values (Bates and Watts 1988) and sensitivities (Pianosi et al. 2016). Fourth, the model predictions are tested with data that has not been used for calibration. Lastly, if confidence in the model has been established, the model can be analyzed with respect to steady states, stability and bifurcations (Kuznetsov 2013; Strogatz 2018) and new experiments and hypotheses can be designed. The corresponding biogeochemical models represent numerous specific microbial and physicochemical regulation mechanisms of nutrient cycling and environmental dependencies at small spatial (μm - cm) and temporal (hourly to daily) scales (Wieder et al. 2015, Table 1). Ultimately however, biogeochemical models aim at predicting system responses at the scale of ecosystems (or field scale, in the case of pollutant turnover). Bridging these scales requires the identification of key mechanisms at the scale at which the empirical studies are conducted and the subsequent development of simpler up-scaled expressions that can be used in ecosystem (or reactive transport, in the case of pollutant turnover) models. These effective descriptions can be used with great confidence, because they *emerge* from mechanistic descriptions of small-scale processes that have been validated in specific experiments. The challenge explored in this thesis is to find a framework in which the relevant mechanistic small-scale descriptions in complex biogeochemical models can be systematically identified and simplified, together with a comprehensible account of the assumptions used and a report on the resulting uncertainties.

Due to the complexity of systems, models are necessarily incomplete representations of reality. The modeling error introduced into process-based models results on the one hand from *conceptual uncertainty* of how to adequately represent biogeochemical processes and on the other hand from *parameter uncertainty*, i.e. the difficulty of assigning values to parameters that describe the process formulations. Orthogonal to questions of experimental design (Keenan et al. 2013) and the associated *measurement uncertainty*, models have to be refined or elaborated over and over again until the modeling process

[1] `https://toolbox.google.com/datasetsearch`

is complete (Getz et al. 2018). Soil biogeochemical modeling has experienced a surge of interest regarding the microbial modeling part of the system over the last few years (Ingwersen et al. 2008; German et al. 2012; Pagel et al. 2014; Wang et al. 2015; Manzoni et al. 2016; Allison 2017). Associated structural and process complexities (Getz et al. 2018) that have been introduced as a result have a large impact on predicted model behavior in response to external environmental drivers, e.g., changes in temperature or carbon input into the system (Allison et al. 2010*a*; Hararuk et al. 2015; Georgiou et al. 2017; Ballantyne IV and Billings 2018). Model structure also has a strong impact on the relevance of parameters that influence the system (Sierra et al. 2015*b*; Vogel et al. 2018). Yet, structural model uncertainty prevails in the microbial component of the system as "the appropriateness of model assumptions cannot be readily assessed because of the scarcity of relevant experimental data" (Baveye et al. 2018). Due to the diversity of the microbial component (Hug et al. 2016) and nonlinear feedback structures in soil systems (Paul 2014), it is tempting to "model bulldozers with quarks" (Goldenfeld and Kadanoff 1999), i.e. it is difficult to choose the right level of model complexity to catch the phenomena of interest. Nonetheless, a fixed model structure is often imposed in biogeochemical modeling. This turns soil biogeochemical models into "gray-box" models (Verghese 2009) with a semi-empirical and highly nonlinear model structure, but many unknown parameters.

A related big challenge in microbially-explicit soil biogeochemical modeling is that "microbial trait parameters are often uncertain, especially when specifying many functional types" (Allison 2017). Most parameter values cannot be measured directly (Fierer et al. 2014). They have to be inferred from data by inverse modeling (Engl et al. 2009; Villaverde and Banga 2014). At a minimum, microbial parameter values related to resource acquisition functions that saturate with increasing carbon concentrations in soil (Tang 2015) have to be evaluated from experimental data. Separate from complex soil biogeochemical models, Strigul et al. 2009 showed that the estimation of microbial parameters from batch experiments in the lab is already a nontrivial problem that requires careful experimental design. Problems associated with kinetic fitting of short-term heterotrophic soil respiration experiments were likewise reported (Wutzler et al. 2012; Weihermüller et al. 2018). In complex models this problem is aggravated as additional microbial trait parameters related to metabolic losses, physiology and substrate preferences combine to determine *emergent* microbial growth rates. Unfortunately, observations are rarely rich enough to allow reliable estimation of model parameters despite the application of sophisticated parameter optimization algorithms (Sierra et al. 2015*a*; Wang et al. 2015; Pagel et al. 2016; Luo et al. 2017). The ill-posed nature of the inverse problem appears to permeate biogeochemical models of multiple complexities. In terms

of parameter *equifinality*, "the whole [might not be] more than the sum of its parts" (Addiscott 2011).

Do natural limits exist in biogeochemical modeling? Soil biogeochemical models are often constructed by combining many independently derived empirical functional relationships. These empirical laws have natural limiting cases: reaction rates can be described by saturating or linear functions of carbon concentrations (Sierra et al. 2012), whole enzyme-substrate networks can be lumped into one rate expression (Tang and Riley 2013), time scales of different system components separate (Manzoni et al. 2016; Kuehn 2016; Kügler 2016; Kügler et al. 2017), etc. Running a biogeochemical model with all possible parametric formulations consistent with a general relationship (Adamson and Morozov 2012) is likely not feasible. Simply deciding on a lower-level description is unsatisfactory due to the high level of structural uncertainty in biogeochemical modeling. Instead of deciding on a lower level of process complexity a priori, "applying [effective modeling] approaches to deriving equations for microbial dynamics in soils could be a promising research direction." (Wieder et al. 2015).

1.3 Modeling of Pollutant Turnover in Soil

Biogeochemical models are complex, but they are powerful tools to model the turnover of pesticides in heterogeneous soil micro-environments by integrating experiments and small-scale mechanistic process descriptions (e.g., Gharasoo et al. 2012; Aslam et al. 2014; Rosenbom et al. 2014; Pagel et al. 2016; Babey et al. 2017). In order to shed light on the processes controlling pesticide degradation in soil at a biogeochemical interface (Totsche et al. 2010), Pagel et al. 2016 integrated genetic information and mechanistic modeling in the development of the PECCAD (Pesticide Degradation Coupled to Carbon Turnover in the Detritusphere) model. PECCAD abstracts specific regulation mechanisms of the degradation of the xenobiotic 4-chloro-2-methylphenoxyacetic acid (MCPA). Microcosm experiments showed that the degradation of MCPA is accelerated in the presence of additional organic substrates (e.g., by litter input; Poll et al. 2010; Saleh et al. 2016). Possible microbial regulation mechanisms were attributed to direct stimulation of a community of specific herbicide degraders or co-metabolic activities of fungi (Ditterich et al. 2013). Thereby, PECCAD explicitly simulates the dynamics of specific microbial degrader populations and accounts for interactions between microbial dynamics and physicochemical processes (sorption and one-dimensional transport of solutes). It captures the complexity of a system response (accelerated MCPA degradation due to litter input) *emerging* from small-scale nonlinear soil processes. MCPA was used as a model compound, because it is predominantly degraded by microorganisms under aerobic conditions and a wealth of untapped biological information on microbial degradation pathways and func-

tional genes involved in MCPA degradation is available (e.g., Bælum et al. 2010; Liu et al. 2013; Nielsen et al. 2013). In a novel modeling approach, the microbial pools in the PECCAD model were set up in correspondence with quantifiable genetic proxies for abundances of total bacteria, fungi and specific pesticide degraders (López-Gutiérrez et al. 2004; Manerkar et al. 2008; Bælum et al. 2008). While integrating soil genomics and mathematical modeling gave detailed insight into MCPA degradation processes and microbial dynamics, the study by Pagel et al. 2016 has pointed to challenges in the calibration and parameterization of the PECCAD model. In particular, the combination of physicochemical, genetic and isotopic data was not able to constrain all model parameters. Parameter uncertainties derived from multi-objective calibration across different experimental treatments were largely attributed to structural deficits in the model. As the most complex biogeochemical model with *equifinality* issues analyzed in this thesis, the PECCAD model serves as an ideal test case to reveal parsimonious, yet mechanistic process formulations that reflect essential physicochemical and microbial regulation mechanisms of pollutant turnover.

1.4 Research Target and Structure of the Thesis

This research starts from the hypothesis that the reported universality of *sloppiness* of predictive multiparameter models (Machta et al. 2013) extends to biogeochemical modeling. If that assumption proved to be true, a unifying framework for model-data integration could be provided that allows to systematically tailor biogeochemical model complexity to the information content of available data and derive effective descriptions of relevant biogeochemical processes. In this thesis, applicability and usefulness of the *sloppy* approach for biogeochemical modeling is assessed for different case studies with focus on parameter identification, structural model limitations and minimal *emergent* model mechanisms that give rise to a predefined system behavior. The insights from four case studies should in the end converge into a set of lessons for how to approach the *equifinality* problem in complex biogeochemical systems.

This thesis opens in Chapter 2 with materials and methods that are necessary to understand implementation details and reproduce the results in the subsequent chapters. In Section 2.1 five soil biogeochemical models of varying complexity from the literature are reviewed (Ingwersen et al. 2008; German et al. 2012; Pagel et al. 2014; Wang et al. 2015; Manzoni et al. 2016). Emphasis is placed on differences in model structures (structural complexity) and functional relationships (process complexity) (Getz et al. 2018). Associated difficulties in obtaining precise parameter estimates from available data by nonlinear least squares regression (Bates and Watts 1988), which partially motivated this thesis, are described. A sensitivity analysis is inevitably tied to the process of, and

problems associated with, model calibration. In Section 2.2 the necessary mathematical tools in order to study the impact of small perturbations on the model performance metric are introduced. Section 2.3 introduces methods of global sensitivity analysis that allow to study how model output is affected by large variations of model parameters. Bayesian ensembles (Stone 2013; Geris et al. 2016) and the Morris method (Pianosi et al. 2016) are outlined that rely on Monte Carlo sampling of the model parameter space. In contrast to these, a recently developed technique from information geometry (the Manifold Boundary Approximation Method (MBAM); Transtrum et al. 2011) is described that only relies on the integration of local sensitivity information (Section 2.2). The final section (Section 2.4) shows how these methods can be used to construct simplified effective models from complex models.

In Chapter 3 the *equifinality* problem associated with the five biogeochemical models described in Section 2.1 is investigated. Locally, all models have *sloppy* sensitivity spectra at their respective published optimal parameter values. In Section 3.3.2 the parameter space of the PECCAD ODE model (Pagel et al. 2014) is analyzed semi-globally in order to determine whether *sloppiness* extends beyond a local phenomenon. In Section 3.3.3 the Monod model for microbial growth (Monod 1949) is used as a simple example to clarify the connection between global sensitivity and information geometry (Section 2.3.3). The chapter ends with a discussion of the implications of *sloppiness* for parameter estimation, parameter space sampling, model selection and model predictions in biogeochemical modeling.

In Chapter 4 problems associated with short-term substrate-induced heterotrophic respiration analysis are investigated. Whether synthetic respiration data provide enough constraining power to identify unique biophysiological parameters of the MEND microbial physiology model (Wang et al. 2014*a*) is analyzed in Section 4.3.1. Problems associated with unique parameter identification are analyzed in Section 4.3.2. Here, *sloppiness* is identified as the underlying cause of non-identifiability in heterotrophic respiration assays.
The issue of parameter non-identifiability is addressed in Section 4.3.3 by constructing reduced models for the respiration assay, using the MBAM. The reduced models from iterative application of the MBAM admit *emergent* parameter combinations that are increasingly more constrained by available data. Section 4.4 then discusses whether the *emergent* parameters still confer biophysically meaningful information about the growth and activity of microorganisms in soil and whether model reduction can and should be used as an additional step in the analysis of heterotrophic respiration data.

In Chapter 5 the over-parameterized PECCAD model (Pagel et al. 2016) is systematically tailored to the information content of data from microcosm experiments by the MBAM. Section 5.3.1 gives details on nontrivial identified limits of the PECCAD model. Different experimental treatments are used as validation data sets to assess the performance of the full and reduced effective models and highlight structural model deficiencies (Section 5.3.1). How model structure changes with the resolution of observations ((i) using information on all data including the dynamics of functional genes; (ii) or bulk biomass along with measurements of dissolved organic carbon, insoluble soil organic matter, total pesticide and heterotrophic respiration; (iii) only with input-output information on total pesticide and CO_2.) is investigated in Section 5.3.1. In Section 5.3.2 the relevant parameters identified by the MBAM are compared with global sensitivity metrics derived from methods that rely on Monte Carlo sampling of the parameter space. In Section 5.4 the regulation mechanisms of MCPA turnover derived by model reduction are compared to conclusions drawn from multi-objective calibration in the original study (Pagel et al. 2016). The predictive power of the PECCAD model is assessed for the first time. The effective parameter combinations that *emerge* from the model reduction steps are compared to reported biogeochemical parameter values from the literature. Limits of the PECCAD model are compared to other commonly used biogeochemical model assumptions.

In Chapter 6 a toy problem is designed that illustrates how to link a specific *emergent* system response function to the underlying biogeochemical mechanisms. Instead of relying on a top-down approach to model pesticide stabilization in soil, the MBAM is used in Section 6.3 to derive a minimal phenomenological model from the complicated, mechanistic PECCAD ODE model. The effective parameters span the degrees of freedom of the system response curve and are linked to resilience theory (Schaeffer et al. 2016). The usefulness and extensions of the approach for up-scaling efforts in biogeochemical modeling are discussed in Section 6.4.

This thesis ends in Chapter 7 with the synthesis of four case studies into fundamental insights on biogeochemical modeling.

Chapter 2

Materials and Methods

2.1 Biogeochemical Model Description

Soil biogeochemical models considered in this thesis can be written as ordinary differential equations (ODE)

$$\frac{dy}{dt} = f(y, \vartheta, t) \tag{2.1}$$

where $y \in \mathbb{R}^M$ is a vector of state variables, $\vartheta \in \mathbb{R}^N$ is a vector of unknown parameters and $t \in \mathbb{R}$ denotes the independent time variable. Given initial conditions, $y(t_0) = y_0$, Eq. 2.1 can be numerically integrated in order to obtain the values of model state variables at discrete time points. Soil biogeochemical systems are typically only partially observed, i.e. observations only exist for subsets or combinations of the total number of model variables M. Moreover, initial conditions y_0 for some model variables have to be estimated from data. For ease of notation, they were included in the parameter vector ϑ.

The prevalent way of fitting biogeochemical models to data is by nonlinear least squares regression (Bates and Watts 1988). Throughout the thesis, this point of view has been adopted and it is assumed that data y_m^D measured at time points t_m are Gaussian distributed with standard deviation σ_m around model predictions $y(\vartheta, t_m)$ of the ODE model (Eq. 2.1). Minimizing the weighted sum of squared errors over all data points M,

$$K(\vartheta) = \frac{1}{2} \sum_{m=1}^{M} \left(\frac{y_m^D - y(\vartheta, t_m)}{\sigma_m} \right)^2 , \tag{2.2}$$

corresponds to a maximum likelihood estimate ($\vartheta = \vartheta^*$) of model parameters,

$$-2logP(y^D|\vartheta) = K(\vartheta) + const\,, \tag{2.3}$$

where the probability density of observing data y^D given parameters ϑ is given by $P(y^D|\vartheta)$.

In synthetic data scenarios, continuous time series data for all model variables that the model fits perfectly at published optimal parameter values ϑ^* were assumed. The associ-

ated cost function is analogously given by

$$K_{cont}(\vartheta) = \frac{1}{2M} \sum_{m=1}^{M} \frac{1}{T} \int_0^T \left(\frac{y_m(\vartheta, t) - y_m(\vartheta^*, t)}{\omega_m} \right)^2 , \tag{2.4}$$

where M is the number of state variables, T is the total simulation time of Eq. 2.1 and the normalization constant ω_m is given by the maximum value of variable y_m across the time interval [0,T] (Gutenkunst et al. 2007).

Five biogeochemical models (Ingwersen et al. 2008; German et al. 2012; Pagel et al. 2014; Wang et al. 2015; Manzoni et al. 2016) were selected from the literature, because they share many features with the majority of soil biogeochemical models (Wieder et al. 2015), but represent different levels of structural and process complexities (Getz et al. 2018). The investigated biogeochemical models promote a mechanistic understanding of soil processes. However, for three out of the five models parameter *equifinality* issues for the fit to data by nonlinear least squares were reported (Sierra et al. 2015*a*; Wang et al. 2015; Pagel et al. 2016). In contrast to the original publications, spatial variations of state variables and transport processes for the models by Ingwersen et al. 2008 and Pagel et al. 2014 were neglected in this study.

2.1.1 A Minimal Microbial Soil Carbon Model

The simplest nonlinear model available from the literature (M=2, N=4; Manzoni and Porporato 2007; German et al. 2012) can be expressed as a system that describes the turnover of soil carbon C_S and microbial biomass C_B

$$\frac{dC_S}{dt} = I(t) - \frac{V_{max} C_S}{K_S + C_S} C_B + k_B C_B \tag{2.5}$$

$$\frac{dC_B}{dt} = Y \frac{V_{max} C_S}{K_S + C_S} C_B - k_B C_B \, . \tag{2.6}$$

The model was developed with the intent to succeed traditional first-order linear soil carbon models (e.g., Jenkinson 1990; Parton et al. 1987) in an earth system modeling context (Wieder et al. 2015; Luo et al. 2016). The assumption that the soil carbon pool decay rate depends on both substrate C_S and catalyst C_B pool sizes leads to significant differences in long term carbon cycle projections in climate change scenarios (Todd-Brown et al. 2013). However, there is no consent as to the correct functional form of nonlinear substrate uptake kinetics in the model (Tang 2015). The regular Michaelis-Menten function used in Eqs. 2.5 and 2.6 assumes that the carbon decay rate saturates at high substrate concentrations. In the regular Michaelis-Menten model, V_{max} is the

maximum growth rate, K_S is the substrate affinity and Y is the microbial carbon use efficiency. Depending on substrate concentration as the rate limiting factor, the rate naturally interpolates between two limiting cases. The rate becomes first-order when substrate concentrations are much smaller than the affinity constant ($C_S \ll K_S$). It becomes constant when soil carbon is in excess ($C_S \gg K_S$). Expressed in terms of parameter limits, the rate becomes linear when maximum growth rate and substrate affinity are both large ($V_{max}, K_S \to \infty$). It becomes constant when the substrate affinity approximates zero ($K_S \approx 0$). Carbon turnover additionally depends on carbon input into the soil $I(t)$, which is time-dependent and often independently estimated from direct measurements, and the first-order cycling rate of microbial biomass k_B.

Sierra et al. 2015*a* discuss the parameter identification problem for Eqs. 2.5 and 2.6 in the context of nonlinear least squares regression (Eq. 2.2). Based on a collinearity index (Soetaert et al. 2010) which summarizes local parameter sensitivity information (Section 2.2), Sierra et al. 2015*a* find that parameters for Eqs. 2.5 and 2.6 cannot be simultaneously identified from commonly available data on respiration fluxes from soil incubation experiments and mass loss data from litter decomposition. Since Eqs. 2.5 and 2.6 are the building blocks of more complex soil carbon turnover models, the identification problem is aggravated in more complex models throughout the biogeochemical modeling literature (Ingwersen et al. 2008; Manzoni et al. 2014; Pagel et al. 2014; Wang et al. 2015). More complex soil biogeochemical models are refined by adding state variables beyond a bulk description of soil carbon and biomass. They are refined with the aim to better capture soil carbon responses with respect to environmental perturbations and to obtain a mechanistic understanding of important biogeochemical regulation mechanisms of carbon cycling. These models operate at smaller temporal (hours to days) and spatial (micrometer to millimeter) scales, but they mostly have to be calibrated on bulk data for soil carbon and microbial biomass.

2.1.2 The Extended NICA Model

Ingwersen et al. 2008, M=10, N=15 (Tables A2 and A3) model small-scale carbon turnover in the detritusphere. The detritusphere is the top millimeter of soil that includes the litter layer and the adjacent soil layer influenced by litter input (Stahr et al. 2016). It is a microbial hotspot in soil with high process rates and nonlinear interactions between constituents (Kuzyakov and Blagodatskaya 2015). Understanding processes in hotspots is relevant for an overall understanding of carbon cycling in soil at larger scales (Totsche et al. 2010; Blagodatsky and Smith 2012). In order to simulate the effects of litter decomposition on carbon turnover in the detritusphere, Ingwersen et al. 2008 partition the soil carbon pool into fractions of different quality. The microbial biomass pool is split

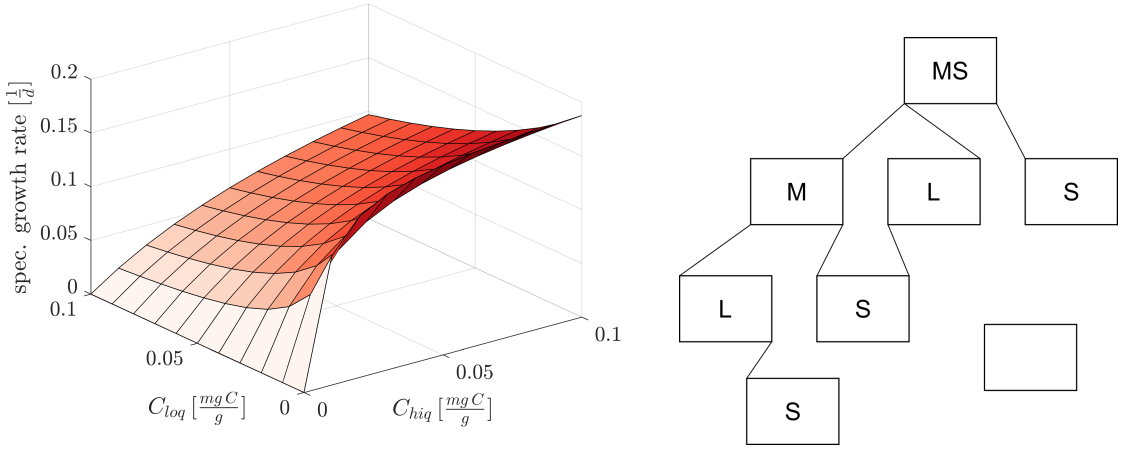
into different functional groups of microorganisms that are thought to reflect bacterial and fungal physiologies. Readily decomposable “initial-stage” dissolved organic carbon (e.g., cellulose and glucose) is exclusively taken up by “initial-stage” microorganisms, whereas more recalcitrant “late-stage” dissolved organic carbon (e.g., interwoven holocellulose and lignin) is solely utilized by “late-stage” decomposers. Microbial physiology is incorporated through empirical functional relationships that describe how microbial growth, maintenance, death and activity depend on carbon concentrations. Ingwersen et al. 2008 assume Monod or regular Michaelis-Menten type kinetics (Strigul et al. 2009; Tang 2015) for growth, maintenance and death terms. The activity of microorganisms is explicitly modeled as a dynamic variable based on the work of Blagodatsky and Richter 1998.

Data on total organic carbon (TOC), dissolved organic carbon (DOC), microbial biomass and ergosterol content, as well as heterotrophic respiration data from a 13C-microcosm experiment (Poll et al. 2008) were used to calibrate the model. Parameter estimation was performed with the Levenberg-Marquardt algorithm for nonlinear least squares problems as implemented in the software UCODE_2005 (Poeter et al. 2006). Uncertainties on the optimized parameter values were not reported (Ingwersen et al. 2008, Table 2).

2.1.3 The PECCAD Model

The carbon model by Pagel et al. 2014, M=12, N=59 (Tables A11 and A12) can be considered as a more refined and elaborate version of the model by Ingwersen et al. 2008. Pagel et al. 2014 model priming effects of litter carbon input on degradation of the herbicide 4-chloro-2-methylphenoxyacetic acid (MCPA) in the detritusphere. The PECCAD (PEsticide degradation Coupled to CArbon turnover in the Detritusphere) model was formulated with the aim to identify important regulation mechanisms of accelerated pesticide degradation in the detritusphere. The hypothesis was that additional supply of resources and energy from litter carbon input into the system can stimulate growth of a community of specific MCPA degraders through improved cosubstrate generation and enzyme production. At the same time, it can enhance production of unspecific enzymes involved in MCPA degradation that are produced by fungi.

In comparison to the model by Ingwersen et al. 2008, the classification of soil microbial populations is refined through the incorporation of a third functional type. Additionally, physicochemical characteristics of soil which control sorption processes for dissolved organic carbon and MCPA are taken into account. Phase concentrations of DOC are related by linear sorption isotherms. Dissolved and sorbed phases of MCPA carbon are related by the Freundlich isotherm. Biokinetic functions used in the PECCAD model are more



(a) Multi-substrate Monod kinetics **(b)** Limiting behavior

Figure 2.1 Multi-substrate Monod kinetics. (a) Synergistic effects of both high-quality (C_{hiq}) and low-quality (C_{loq}) carbon concentrations on the specific growth rate of bacteria. (b) Model hierarchy for the multi-substrate Monod model (MS). Monod (M), linear (L) and saturating (S) kinetics correspond to distinct parameter limits as described in the main text. No reaction (empty square) is a special case of all sub-types.

elaborate, because they simulate the dependence of microbial growth, maintenance and death rates in terms of multi-substrate Monod kinetics (Lendenmann and Egli 1998). In contrast to single substrate Monod kinetics, the specific growth rate of, e.g., bacteria (μ_B) is a function of both high-quality (C_{hiq}) and low quality (C_{loq}) carbon:

$$\mu_{B,i} = \frac{\mu_{max-B} k_{B,i} C_i}{\mu_{max-B} + k_{B,hiq} C_{hiq} + k_{B,loq} C_{loq}}\,, \quad i \in \{hiq, loq\}\,. \tag{2.7}$$

Multi-substrate Monod kinetics simulate the synergistic effect of different carbon sources on the specific growth rate (Fig. 2.1a). The hierarchical structure of the growth model is depicted in Fig. 2.1b. Different parameter limits interpolate between Monod-type (M), linear (L) and saturating (S) kinetics. E.g., for $i = loq$, Monod kinetics are recovered when the substrate affinity of bacteria to C_{hiq} approximates zero ($k_{B,Hiq} \approx 0$). Linear kinetics correspond to the parameter limit in which the maximum specific growth rate approaches infinity ($\mu_{max-B} \to \infty$). Saturating kinetics are retrieved when the magnitude of the low-quality substrate affinity coefficient is much larger than carbon concentrations in the system ($k_{B,loq} \to \infty$). The price to pay for increased realism or flexibility of the multi-substrate Monod formulation compared to single Monod kinetics is certainly that an additional substrate affinity coefficient $k_{BP,i}$ per biokinetic function has to be estimated from data. Although the PECCAD model is structurally similar to the model by

Ingwersen et al. 2008, the number of parameters almost quadrupled as a result of the more complex process formulations.

The model was calibrated on multi-modal, genomic soil carbon data from a series of microcosm experiments (Pagel et al. 2016). The experimental design is described in detail in Pagel et al. 2016. A brief account of the experimental treatments is given in Section 5.2. Data included MCPA concentrations, extractable DOC, TOC, microbial biomass, genetic abundances of bacteria (*16S rRNA* genes), fungi (ITS fragments) and specific MCPA degraders (*tfdA* genes) as well as isotopically labeled compounds. A multi-objective Pareto analysis (the AMALGAM method; A Multi-ALgorithm, Genetically Adaptive Multiobjective method; Vrugt and Robinson 2007) was used to find pareto-optimal parameter sets for the different experimental treatments. However, Pagel et al. 2016 concluded that "the available data might not have been sufficient to simultaneously identify the large number of parameters associated with the high complexity of the PECCAD model". Indeed, the range of Pareto optima for 26 out of 59 estimated biokinetic parameters approximated their respective prior ranges after optimization (Pagel et al. 2016, Table 2).

2.1.4 The MEND Model

The microbial-enzyme-mediated decomposition (MEND, Tables A5 and A6) model (Wang et al. 2013, 2015, M=10, N=19) adds structural complexity to carbon cycling in comparison to the previously described soil biogeochemical models (German et al. 2012; Ingwersen et al. 2008; Pagel et al. 2014). It was developed in order to better predict climate feedbacks from microbially-mediated carbon cycling on an ecosystem scale. Carbon decomposition in the MEND model is controlled by different extracellular enzymes. A particulate organic carbon pool including lignocellulose-like compounds is broken down by ligninases and cellulases. A second particulate organic carbon pool containing starch-like compounds is degraded by amylases. Extracellular enzymes are solely produced by active biomass and shifts in maintenance requirements of the total biomass allocated to enzyme production are accounted for. A contemporary model for the physiological activity state of microorganisms (Wang et al. 2014*a*) was incorporated. The dormancy model is coupled to microbial maintenance requirements, instead of carbon saturation levels. It remedies short-comings of the original formulation by Blagodatsky and Richter 1998 at low substrate concentrations in soil that was used by Ingwersen et al. 2008 and Pagel et al. 2014. Otherwise, the MEND model shares structural elements of carbon cycling with these two models. The complexity of functional relations is similar to the model by Ingwersen et al. 2008, because biokinetic functions are expressed in terms of single Monod kinetics.

The MEND model was calibrated on long-term carbon decomposition data from laboratory incubations of four soils across different climatic regions. Data included total heterotrophic soil respiration, microbial biomass, DOC and isotopically labeled carbon fractions for two different experimental treatments (glucose and starch amendments). The SCEUA (Shuffled Complex Evolution at University of Arizona, Duan et al. 1992) algorithm in unison with a Critical Objective Function Index (COFI, Wang et al. 2015) was used to estimate 19 biogeochemical model parameters and report parameter uncertainties. The authors conclusion echoes the sentiment of Pagel et al. 2016: "We hope that these parameter values could provide documented references, although with large uncertainty, for relating parameters to soil properties and/or climatic conditions. Nevertheless, more laboratory- and field-scale data sets are needed to constrain model parameterization and associated uncertainty.". The intrinsic microbial carbon use efficiency (true growth yield) is a fundamental parameter for ecological models that is difficult to measure in situ (Sinsabaugh et al. 2013). Its temperature sensitivity was in particular not well constrained by the available soil incubation data (Wang et al. 2015, Figure 4).

2.1.5 A Trait-Based Microbial Soil Carbon Model

The model by Manzoni et al. 2016, M=7, N=24 (Tables A8 and A9) differs conceptually from the other four soil biogeochemical models described in this thesis ((Ingwersen et al. 2008; German et al. 2012; Pagel et al. 2014; Wang et al. 2015)). It is purely a theoretical study aimed at understanding how microbial physiological traits, ecological strategies and transport processes influence biogeochemical processes in varying soil moisture regimes across different scales. The soil moisture balance that describes variations of the environmental control variable is formulated at the operating scale of the model (carbon pools are expressed in units of $g\,C/m^3$, varying at a daily time scale). The focal level of model observations (Getz et al. 2018) is to be interpreted as a cubic meter of soil from which respiration fluxes are measured. The heterotrophic respiration flux represents an emergent large scale pattern that results from small-scale soil processes that are explicitly modeled. Small-scale processes include different microbial response strategies to water stress (the regulation of osmolyte production, drought avoidance by dormancy and increased extracellular enzyme production) and diffusion. Mass transfer coefficients for the diffusion of enzymes and soluble carbon compounds were derived for characteristic distances of $10^{-5} - 10^{-3}$ m. Empirical laws that are based on lab experiments describe how solute diffusivity (Hamamoto et al. 2010) or osmolyte production (Dötsch et al. 2008) depend on volumetric soil moisture. These process descriptions represent model refinements below the focal level of typical soil carbon models. Model structure deviates from other soil carbon models accordingly, because pools for intracellullar osmolyte produc-

tion were added. Moreover, biokinetic functions for carbon uptake by microorganisms are linear ($\propto V_{max} C_S C_B$), rather than saturating functions of substrate concentrations.

The model was not fit to data. Instead, it is part of a pattern-oriented (Grimm and Railsback 2012) modeling effort to derive microbial response patterns to environmental control factors directly from the underlying micron-scale processes (Allison 2017). The model is used as a mechanistic basis for the on-going development of a hierarchy of process models of different complexity that try to capture "Type 1" and "Type 2" microbial responses to rewetting of soil (Brangari et al. 2018).

2.2 Local Sensitivity Analysis

Local sensitivity analysis studies the effect of small perturbations of model parameters on the model output. When model calibration is considered, the model output metric is given by the fit of the model to given data. Local sensitivity of the model fit to variations in parameter values in the vicinity of the best-fit point ϑ^* is given by the quadratic term in the Taylor expansion of the cost function $K(\vartheta)$ (Eq. 2.2):

$$K(\vartheta) = K(\vartheta^*) + \nabla K(\vartheta^*)^T \Delta\vartheta + \frac{1}{2} \Delta\vartheta^T \, H \, \Delta\vartheta + \mathcal{O}(|\Delta\vartheta|^3) \, . \tag{2.8}$$

The best fit ϑ^* corresponds to the local minimum of the cost function at which the gradient $\nabla K(\vartheta^*)$ is zero. Here, $\Delta\vartheta = (\vartheta - \vartheta^*)$ and T denotes the transpose of a matrix, $M_{ij}^T = M_{ji}$, for which the i-th row, j-th column element of the transpose is the j-th row, i-th column element of the original matrix M. The Hessian H is an $N \times N$ matrix of second-order partial derivatives of the scalar, multi-variate cost function evaluated at the best-fit point. It describes the local curvature of the cost function away from the best fit:

$$H_{ij} = \frac{\partial^2 K}{\partial\vartheta_i \partial\vartheta_j}\Big|_{\vartheta=\vartheta^*} \, . \tag{2.9}$$

Since H is symmetric, it can be factorized in terms of its eigendecomposition as $H = Q\Lambda Q^T$, where Q is an orthogonal matrix, $Q^T Q = 1$, the N columns of which are normalized eigenvectors of H and Λ is a matrix with the corresponding positive eigenvalues along the diagonal. The deviation of the cost from its minimum value can then be written as

$$\Delta K(\vartheta) = K(\vartheta) - K(\vartheta^*) \approx \frac{1}{2} \Delta\vartheta^T \, Q\Lambda Q^T \, \Delta\vartheta = \frac{1}{2} \xi^T \Lambda \, \xi = \frac{1}{2} \sum_{n=1}^{N} \lambda_n \xi_n^2 \, , \tag{2.10}$$

which shows that there is a change of basis $\xi = Q^T \Delta\vartheta$ such that the matrix of the quadratic form is diagonal. The transformed vectors ξ and hence the normalized Hessian

eigenvectors define the principal axes of an N-dimensional ellipsoid that is generally rotated with respect to the bare parameter axes. For example in two dimensions, N=2, a cost function isocurve (Eq. 2.10) is given by

$$2\Delta K(\vartheta) = \lambda_1 \xi_1^2 + \lambda_2 \xi_2^2 = \frac{\xi_1^2}{(\frac{1}{\sqrt{\lambda_1}})^2} + \frac{\xi_2^2}{(\frac{1}{\sqrt{\lambda_2}})^2}\,, \tag{2.11}$$

where the shape parameters that define the major and minor axis of this ellipse are scaled by the inverse of the square root of the Hessian eigenvalues λ_1, λ_2. Projections of the ellipse onto the i-th bare parameter axis are proportional to $\sqrt{H_{ii}^{-1}}$. Intersections with the i-th axis are given by $\sqrt{1/H_{ii}}$.

In terms of the model residuals,

$$r_m(\vartheta) = \frac{y_m^D - y(\vartheta, t_m)}{\sigma_m}\,, \tag{2.12}$$

second derivatives of the cost function can be expressed as

$$H_{ij} = \frac{\partial^2 K}{\partial\vartheta_i \partial\vartheta_j}\Big|_{\vartheta=\vartheta^*} = \sum_m \frac{\partial r_m}{\partial\vartheta_i}\frac{\partial r_m}{\partial\vartheta_j}\Big|_{\vartheta=\vartheta^*} + \sum_m r_m \frac{\partial^2 r_m}{\partial\vartheta_i\partial\vartheta_j}\Big|_{\vartheta=\vartheta^*}\,. \tag{2.13}$$

A valid assumption close to the best-fit point is that residuals are small ($r_m \approx 0$) and the second term in Eq. 2.13 can be neglected. An approximation of the Hessian is then given in terms of the scaled Jacobian matrix J_σ:

$$H_{ij} \approx \sum_m \frac{\partial r_m}{\partial\vartheta_i}\frac{\partial r_m}{\partial\vartheta_j} = \sum_m \frac{1}{\sigma_m^2}\frac{\partial y_m}{\partial\vartheta_i}\frac{\partial y_m}{\partial\vartheta_j} \equiv J_\sigma^T J_\sigma\,. \tag{2.14}$$

Entries of the $M \times N$ Jacobian matrix J are partial derivatives of M model outputs at selected measurement time points with respect to N model parameters, $J_{mi} = \partial y_m/\partial\vartheta_i$. In practice, local sensitivity coefficients $S_{ij} \equiv \partial y_i/\partial\vartheta_j$ can, e.g., be obtained via finite differences or the direct differential method. A finite difference approximation is given by

$$S_{ij} = \frac{\partial y_i}{\partial\vartheta_j} \approx \frac{y_i(\vartheta + \Delta\vartheta_j) - y_i(\vartheta)}{\Delta\vartheta_j}\,, \tag{2.15}$$

but the solution is typically not well-behaved numerically due to the strong dependence on the chosen step size $\Delta\vartheta_j$. For an ODE model (Eq. 2.1), the sensitivity coefficients also satisfy the equation for the total differential

$$\frac{\partial S_{ij}}{\partial t} = \frac{\partial}{\partial t}\frac{\partial y_i}{\partial\vartheta_j} = \frac{\partial}{\partial\vartheta_j}\frac{\partial y_i}{\partial t} = \frac{\partial f_i(y,\vartheta,t)}{\partial\vartheta_j} = \frac{\partial f_i}{\partial\vartheta_j} + \sum_j \frac{\partial f_i}{\partial y_j} S_{ij}\,, \quad S_{ij}(0) = 0\,, \tag{2.16}$$

which is linear, but has to be solved simultaneously with Eq. 2.1 for every parameter. More advanced techniques use automatic differentiation (e.g., Revels et al. 2016) or adjoint methods (in a biological context described in Zi 2011). In order to compute an approximation of the Hessian (Eq. 2.14), the m-th row of the Jacobian matrix must be multiplied by one over the corresponding measurement error σ_m.

When the Hessian matrix H is invertible, the covariance matrix is given by

$$Cov(\vartheta) = 2H^{-1} \tag{2.17}$$

and can be used to derive asymptotic confidence intervals for model parameters

$$\Sigma_i^{\pm} = \vartheta^* \pm \sqrt{\chi^2(\alpha, df) \cdot Cov_{ii}}\,, \tag{2.18}$$

where $\chi^2(\alpha, df)$ is the α-quantile of the χ^2-probability distribution with df degrees of freedom and Cov_{ii} is the corresponding diagonal entry of the covariance matrix. For $df = 1$, Eq. 2.18 gives pointwise confidence intervals known as the standard errors. Simultaneous confidence intervals that correspond to the projection of the N-dimensional ellipsoid onto the i-th bare parameter axis are obtained when $df = N$, i.e. the degrees of freedom are equal to the number of parameters (Tönsing et al. 2014). Asymptotic confidence intervals are only an approximation of the true parameter uncertainty when the observables depend nonlinearly on model parameters. Implied is a limit of infinite data which is approached if the number of data points is large compared to the number of parameters (or the measurement error is small). The Cramer-Rao inequality places a lower bound on the covariance matrix

$$Cov(\vartheta) \geq \frac{H^{-1}}{n} \tag{2.19}$$

stating that if the Hessian is nearly singular, it will take a large number of experimental repetitions n in order to obtain precise parameter estimates.

2.3 Global Sensitivity Analysis

2.3.1 Method of Morris

The Morris method (Morris 1991), also called the Elementary Effect Test (EET, Saltelli et al. 2008; Pianosi et al. 2016) for global sensitivity analysis is a derivative-based OAT (One-step-At-a-Time) method that generates two sensitivity measures for each model parameter: μ^*, the Morris mean and σ^*, the standard deviation.

In an OAT design, each parameter is locally varied for a point in parameter space, while the other parameters are fixed to a nominal value, and the change in model output is recorded. If K denotes the cost function, the finite difference

$$EE_i = \frac{K(\vartheta_1, ..., \vartheta_{i-1}, \vartheta_i + \Delta\vartheta_i, \vartheta_{i+1}, ..., \vartheta_n) - K(\vartheta_1, ..., \vartheta_n)}{\Delta\vartheta_i} \tag{2.20}$$

is called the Elementary Effect (EE) for parameter i, $i \in \{1, ..., N\}$. The EE is the ratio between the variation in the cost metric due to local variation of the parameter and the variation in the parameter itself. In order to obtain a global sensitivity measure, the absolute values of r different EEs for each parameter are computed and averaged to the Morris mean

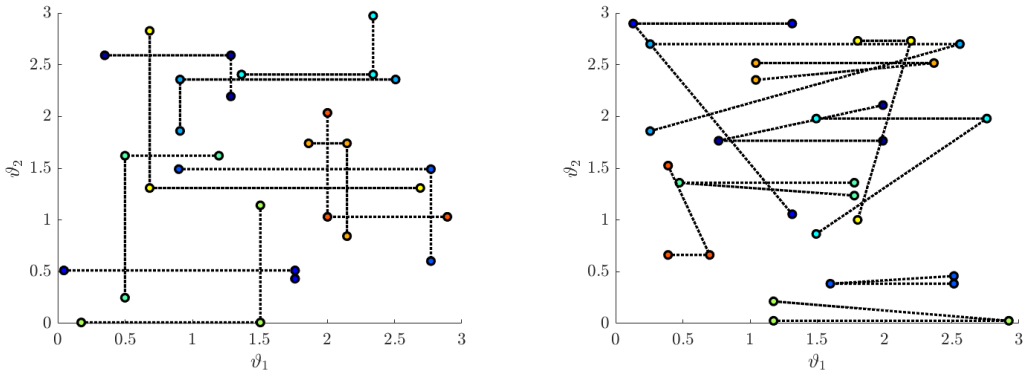
$$\mu_i^* = \frac{1}{r} \sum_{j=1}^{r} |EE_i|^j \,, \tag{2.21}$$

with standard deviation

$$\sigma_i^* = \sqrt{\frac{1}{r-1} \sum_{j=1}^{r} \left(EE_i^j - \mu_i^*\right)^2} \,. \tag{2.22}$$

State-of-the-art versions of the Morris method mainly differ in the sampling strategies used to select initial and consecutive points in parameter space for variation (Pianosi et al. 2016). Examples of elementary LH-OAT sampling strategies for two-dimensional parameter space are shown in Fig. 2.2.

The Morris method (Morris 1991) builds r trajectories in parameter space. Along a trajectory each parameter ϑ_i is consecutively increased or decreased by a fixed step size $\Delta\vartheta_i$. One EE per parameter can be estimated per trajectory. The starting points (also called support points or stars) for the r trajectories are selected randomly over a uniform grid. Van Griensven et al. 2006 improved part of the sampling strategy by generating starting points of the trajectories from a Latin Hypercube Sample (LHS, McKay et al. 1979). Their LH-OAT approach was used by Pagel et al. 2014 in order to assess global sensitivity of parameters in the PECCAD model (Section 2.1). A superior OAT-design for estimation of EEs is the radial-based design (Campolongo et al. 2011). In a radial-based design, parameters along a trajectory are not varied consecutively. Instead, ϑ_i is systematically varied starting from the same support point in parameter space. Parameter variations $\Delta\vartheta_i$ are generally not taken to be the same for each ϑ_i. However, the radial-design implemented in the MATLAB SAFE toolbox (Pianosi et al. 2015) that is used in this study (Fig. 2.2b) differs from the one benchmarked in Campolongo et al. 2011. The design is not necessarily best practice. Samples for the $\Delta\vartheta_i$ are not based on Sobol sequences, but are also generated from a random LHS and the trajectories are not selected as to maxi-



(a) LH-OAT trajectory design **(b)** LH-OAT radial design

Figure 2.2 Examples of elementary parameter space sampling designs. Colored circles denote support and end points of different samples.

mize their dispersion in parameter space (Campolongo et al. 2007; Link et al. 2018).

Due to the comparatively low computational cost of $r \cdot (N+1)$ model evaluations, variants of the Morris method are often used for screening purposes of model input variability. Typically, parameters are grouped into three categories depending on their Morris mean and standard deviation $\{\mu_i^*, \sigma_i^*\}$ (Iooss and Lemaître 2015). The larger μ_i^*, the larger the effect of the i-th parameter on the cost metric. σ_i^* is a measure for nonlinearity or interaction effects for the i-th parameter. If σ_i^* is small, the EEs for the i-th parameter do not vary significantly over support points in parameter space. If the effect of a small perturbation of a parameter is the same everywhere, a linear relationship between parameter and model performance metric is likely. A parameter with large σ_i^* will have non-linear or interaction effects. Different sets of Morris mean and standard deviation hence correspond to parameters that have negligible effect on the cost metric (both μ_i^*, σ_i^* small), those that have a linear effect ($\mu_i^* > \sigma_i^*$, with σ_i^* small) and those with significant interaction effects ($\mu_i^* < \sigma_i^*$, with both μ_i^*, σ_i^* large).

2.3.2 Bayesian Ensembles

Bayesian inference consists of conditioning a prior probability distribution of model parameters on the data (Stone 2013; Geris et al. 2016). Samples from the posterior distribution of parameters resulting from this inference process can be analyzed by descriptive statistics and used to report uncertainties on model parameters and predictions beyond a linearized description (Eq. 2.18). Mathematically, Bayes theorem for conditional prob-

abilities is stated as:

$$P(\vartheta|y^D) = \frac{P(y^D|\vartheta)P(\vartheta)}{P(y^D)} \propto P(y^D|\vartheta)P(\vartheta)\,. \tag{2.23}$$

$P(y^D|\vartheta)$ is the likelihood that the model will produce the data y^D given parameters ϑ (Eq. 2.3). $P(\vartheta)$ is the prior probability distribution of model parameters. Together, likelihood and prior encode the belief of the modeler about observations of the biogeochemical system. $P(y^D)$ is the evidence for the data. The evidence is an unimportant normalization constant in global sensitivity applications.

For a global exploration of parameter space, the objective of a Bayesian Markov Chain Monte Carlo (MCMC) approach is to explore the full nonlinear cost surface of the model. (Bayesian) MCMC algorithms comprise a class of numerical routines that sample from the (posterior) probability distribution $P(y^D|\vartheta)$. The underlying idea is to maximize information about the posterior distribution by sampling densely in areas of parameter space corresponding to good fits to the data and sparsely elsewhere (Mannakee et al. 2016). Ideally, the Markov chain has the posterior as its unique stationary distribution (Tierney 1994), i.e. it eventually converges to the true posterior distribution.

For the Metropolis-Hastings algorithm (Chib and Greenberg 1995), a single Markov chain is started from the presumed best fit $\vartheta_0 = \vartheta^*$. A candidate for the next parameter sample ϑ_c is randomly chosen and accepted according to the following criterion that only depends on the current sample value ϑ_i:

$$\vartheta_{i+1} = \begin{cases} \vartheta_c, & \text{with probability } \alpha \\ \vartheta_i, & \text{with probability } 1-\alpha \end{cases} \tag{2.24}$$

The acceptance ratio α is based on a comparison of the posterior probability for the candidate and current parameter vector:

$$\alpha = \min\left(\frac{P(\vartheta_c|y^D)}{P(\vartheta_i|y^D)},\ 1\right)\,. \tag{2.25}$$

Moves to new points in parameter space ϑ_{i+1} are accepted when the posterior probability of the candidate is higher. If the posterior probability is lower, the move is sometimes rejected. Since the probability of rejection depends on the relative drop in posterior probability, the Markov chain tends to sample in high likelihood regions.

For sloppy models, Gutenkunst 2007 suggests to sample the candidate parameter vector from a multivariate Gaussian distribution, the inverse covariance matrix of which is the Hessian matrix (Eq. 2.17). The acceptance probability that satisfies detailed balance then reads:

$$\alpha = \min\left(\frac{P(\vartheta_c|y^D)}{P(\vartheta_i|y^D)} \cdot \frac{|H_c| \exp\left(-\frac{1}{2}(\vartheta_c - \vartheta_i)^T H_c\,(\vartheta_c - \vartheta_i)\right)}{|H_i| \exp\left(-\frac{1}{2}(\vartheta_c - \vartheta_i)^T H_i\,(\vartheta_c - \vartheta_i)\right)},\ 1\right). \tag{2.26}$$

Here, H_c and H_i are the Hessian matrices calculated at the candidate and current sample points. $|H| \equiv \det H$ is the determinant of H. The idea behind the importance sampling scheme is to avoid steps in stiff directions in parameter space that would yield low acceptance probabilities.

A comparison of the marginal posterior distributions of model parameters with the marginal prior parameter distributions can be used to assess the learning effect of the Bayesian inference process. Narrower marginal posterior distributions compared to the priors indicate good identifiability of model parameters. Large shifts in the maximum a posteriori probability estimate (MAP), i.e. the mode of the marginal posterior distribution, compared to the MAP of the priors, should lead to a check of prior information. The highest posterior density region is the set of most probable parameter values that constitute $100 \cdot (1-\alpha)\%$ of the posterior mass. For a given α, the integral

$$1 - \alpha = \int_{\vartheta:\, p(\vartheta|y^D) > p^{**}} p(\vartheta|y^D) d\vartheta \tag{2.27}$$

defines the set of highest posterior densities $C_\alpha(D) \equiv \{\vartheta :\ p(\vartheta|y^D) \geq p^{**}\}$.

2.3.3 Information Geometry

The field of information geometry combines methods of differential geometry and information theory (Amari and Nagaoka 2007; Ay et al. 2017). The central approach of information geometry is to introduce a notion of distance between probability distributions (Caticha 2015). Many measures of statistical distances between probability distributions have been defined (Burbea and Rao 1982). Some important examples include special cases of f-divergences (Kullback-Leibler (KL) divergence, Hellinger distance, total variation distance), the closely related Bhattacharyya distance, the Kolmogorov-Smirnov statistic for univariate probability distributions or the Kantorovich metric. The Fisher-Information Metric (FIM) is a special case. Endowed with the FIM, a statistical manifold[1]

[1]Loosely speaking, an N-dimensional statistical manifold is "a collection of points that are connected to each other in a smooth fashion such that the neighborhood of each point looks like the neighborhood

corresponds to a Riemannian manifold[2], thus allowing the use of differential geometry (Do Carmo 2016; Abbena et al. 2017).

Geometry is an intrinsic property of statistical manifolds. To see this, consider the expected value of the infinitesimal distance between two probability distributions $P(y^D|\vartheta)$ and $P(y^D|\vartheta + \Delta\vartheta)$, also called the expected value of the score in information theory - a statistical measure of the sensitivity of the log-likelihood function (Eq. 2.3) with respect to infinitesimal parameter variations[3]:

$$\langle \Delta \rangle = \int dy^D P(y^D|\vartheta) \frac{\partial \log P(y^D|\vartheta)}{\partial \vartheta^\alpha} d\vartheta^\alpha = d\vartheta^\alpha \frac{\partial}{\partial \vartheta^\alpha} \underbrace{\int dy^D P(y^D|\vartheta)}_{=1} = 0\,. \tag{2.28}$$

Since the expected value vanishes, the correct measure of distinguishability is the second moment of the score - also called the infinitesimal KL divergence:

$$\langle \Delta^2 \rangle = \int dy^D P(y^D|\vartheta) \frac{\partial \log P(y^D|\vartheta)}{\partial \vartheta^\alpha} \frac{\partial \log P(y^D|\vartheta)}{\partial \vartheta^\beta} d\vartheta^\alpha d\vartheta^\beta\,. \tag{2.29}$$

Equation 2.29 suggests to define the FIM as

$$g_{\alpha\beta} \equiv \int dy^D P(y^D|\vartheta) \frac{\partial \log P(y^D|\vartheta)}{\partial \vartheta^\alpha} \frac{\partial \log P(y^D|\vartheta)}{\partial \vartheta^\beta}\,, \tag{2.30}$$

so that

$$dl^2 \equiv \langle \Delta^2 \rangle = g_{\alpha\beta} d\vartheta^\alpha d\vartheta^\beta \tag{2.31}$$

can be identified as the first fundamental form of differential geometry (the generalized Phythagorean formula). A small infinitesimal line element dl means that the relative distance between nearby probability distributions is small and model predictions from ϑ and $\vartheta + \Delta\vartheta$ will be difficult to distinguish. The FIM thus measures how distinguishable two nearby probability distributions are from given data y^D. The FIM is symmetric, positive definite and transforms like a covariant rank-2 tensor. Its natural interpretation

of an N-dimensional Cartesian space" (Hassani 2013). For statistical manifolds, each point, labeled by coordinates ϑ, represents a probability distribution $P(y^D|\vartheta)$.

[2]A Riemannian manifold is a manifold equipped with a Riemannian metric. A Riemannian metric defines an inner product on the tangent space of each point on the manifold. It allows to locally calculate geometric notions such as angles or areas. However, comparing vectors in nearby tangent spaces in order to do calculus, e.g to compute gradients of functions and divergences of vector fields, requires a metric connection. Concepts inherent to curved surfaces such as the Riemann curvature tensor, parallel transport and geodesics can be expressed in terms of Christoffel symbols that are related to the metric connections.

[3]In differential geometry it is custom to adopt Einstein's summation convention: repeated indices (in this case α) are summed over when they appear as super- and subscripts in an expression, each index can appear at most twice in any term and each term must contain identical non-repeated indices.

as a Riemannian metric is the cornerstone of the field of information geometry[4]. Under coordinate transformations, $d\vartheta' = \frac{\partial \vartheta'}{\partial \vartheta} d\vartheta$, the FIM transforms as

$$g_{\alpha'\beta'} = \frac{\partial \vartheta^{\alpha'}}{\partial \vartheta^{\alpha}} \frac{\partial \vartheta^{\beta'}}{\partial \vartheta^{\beta}} g_{\alpha\beta} \,, \tag{2.32}$$

leading to

$$dl'^2 = g_{\alpha'\beta'} \, d\vartheta^{\alpha'} d\vartheta^{\beta'} \,, \tag{2.33}$$

which shows that the distance dl^2 is an invariant scalar under the coordinate transformation. Coordinates do not have an immediate metrical meaning (Ricci and Levi-Civita 1900). As a result, the statistical manifold and its geometrical properties are the same, regardless of how the model is parameterized.

The FIM is intrinsic to statistical manifolds, but it depends on the underlying statistical data model $P(y^D|\vartheta)$. For nonlinear least squares problems (Eq. 2.3) it follows from Eq. 2.30 that the metric is simply given by (Transtrum et al. 2011)

$$g_{\alpha\beta}(\vartheta) = (J_\sigma^T J)_{\alpha\beta} = \sum_m \frac{1}{\sigma_m^2} \frac{\partial y_m}{\partial \vartheta_\alpha} \frac{\partial y_m}{\partial \vartheta_\beta} \,. \tag{2.34}$$

The FIM in this case is just the approximate Hessian matrix of the inference problem (Eq. 2.14). With an appropriate reparameterization (Eq. 2.32), it would be possible to transform eigenvalues of the Hessian and alter the shape of the local cost surface in parameter space (Section 2.2). In the geometric interpretation however, the scaled Jacobian matrix

$$(J_\sigma)_{m\alpha} = \frac{1}{\sigma_m} \frac{\partial y_m}{\partial \vartheta_\alpha} \tag{2.35}$$

defines a mapping between N-dimensional parameter space with axes labeled by parameters ϑ_i, $i \in N$ and M-dimensional (scaled) data space with axes labeled by y_m/σ_m, $m \in M$. The FIM acts as a Riemannian metric for the model prediction manifold (the set of all possible model predictions y_m labeled by parameters ϑ) which is embedded in the Euclidean data space (for an illustration, see Fig. 2.3). In general, for a model with N structurally identifiable parameters (Villaverde et al. 2016) that is fit to M data points, the model manifold is the N-dimensional surface embedded in $\mathbb{R}^M$ formed by varying model parameters over predefined ranges (Mannakee et al. 2016). The model manifold is a global representation of the model and its properties are invariant under reparameterizations.

[4] In fact it can be shown that the FIM is, up to a constant scale factor, the unique Riemannian metric of a statistical manifold (Cencov 2000). Other derivations of the metric, e.g., from the relative entropy are available in Caticha 2015.

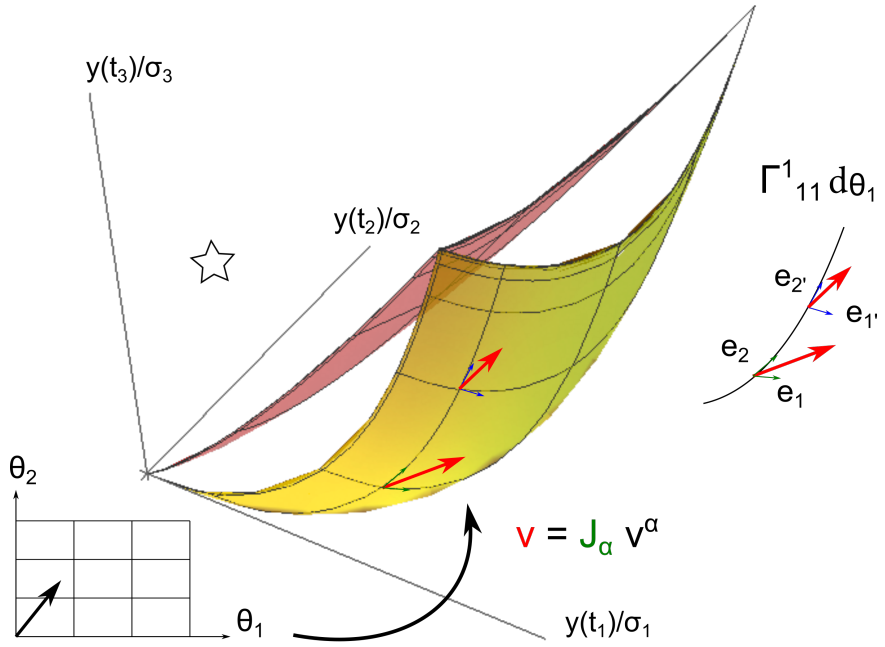


Figure 2.3 Model manifolds of the exponential model $y(\vartheta, t) = e^{-\vartheta_1 t} + e^{-\vartheta_2 t}$. The Euclidean embedding space is spanned by model output recorded at three time points t_s. The best fit is given by the point on the manifold with closest Euclidean distance to the data point (black star). The parameters ϑ_1, ϑ_2 define curvilinear coordinates on the model manifold that correspond to a rectangular grid in parameter space. Simpler models can be identified at the boundaries of the manifold ($\vartheta_1 = 0$, $\vartheta_2 = \infty$ or $\vartheta_1 = \vartheta_2$). Practical identifiability of model parameters changes with sampling times (yellow: $t_s = \{1/3, 1, 3\}$, red: $t_s = \{1/3, 1.5/3, 2/3\}$). The column vectors of the Jacobian J locally span the tangent plane on the manifold. The Christoffel symbol Γ determines how basis vectors change along a path on the manifold.

The data is a single point in $\mathbb{R}^M$ that lies off of the manifold in the case that the model cannot fit the data perfectly. Under the coordinate transformation $\vartheta \to \vartheta' = y$, the FIM is proportional to the ordinary Euclidean metric

$$g_{\alpha\beta}(y) = \frac{1}{\sigma^2}\delta_{\alpha\beta}\,, \tag{2.36}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta[5]. Distance from any point on the manifold through the embedding space to the data point is hence proportional to the nonlinear least squares cost function (Eq. 2.2). Given a realization of the model parameter vector ϑ^α, this distance determines the goodness-of-fit to data.

Distances on the model manifold between close-by statistical models, on the other hand, are given by Eq. 2.35. Through the definition of the metric, distances are expressed in units of standard deviations σ_m of the data. The measurement error is the "yardstick" of nonlinear least squares models. If any dimension of the model manifold is thinner than a standard deviation, model predictions will quantifiably not be distinguishable from data (White et al. 2016). Conversely, the corresponding nonlinear parameter combinations cannot be inferred from available data.

The Jacobian (Eq. 2.35) locally maps an N-dimensional unit cube from parameter space onto an n-parallelotope on the model manifold with volume proportional to $\sqrt{|g|}$. Eigenvalues of the metric tensor determine the degree of anisotropy of the mapping. Globally, Transtrum et al. 2010 empirically showed that model manifold widths approximately correspond to the square root of FIM eigenvalues. Boundaries of the model manifold correspond to points where the FIM becomes singular. This occurs when the column vectors of the Jacobian are linearly dependent and at least one row/column of the metric tensor can be written as a linear combination of the other rows/columns. In its interpretation of the Hessian matrix (Eq. 2.13), a singular metric implies that the model residuals become insensitive to changes in the corresponding (nonlinear) combination of model parameters that define the manifold boundary.

Information geometry enables a global analysis of model parameter sensitivities by exploring geometric features (widths and boundaries) of the model manifold by geodesics. Locally, the column vectors of the Jacobian mapping (Eq. 2.35) define a tangent plane at

[5]The Kronecker delta symbol $\delta_{\alpha\beta}$ that determines the entries of the metric tensor equals 1 if $\alpha = \beta$, and 0 otherwise.

each point on the model manifold spanned by the basis vectors

$$e_\alpha^m = \frac{\partial r_m}{\partial \vartheta_\alpha}\,. \tag{2.37}$$

A geodesic describes parallel transport, i.e. it is the special curve on the model manifold whose tangent vectors remain parallel if transported along it:

$$\frac{\partial e_\alpha^m}{\partial \vartheta_\nu} = \Gamma_{\mu\nu}^\rho e_\rho^m \neq 0\,, \tag{2.38}$$

where the Christoffel symbols Γ, that connect different tangent spaces, can be directly expressed by the metric coefficients:

$$\Gamma_{\mu\nu}^\alpha = \frac{1}{2} g^{\alpha m}\left(\frac{\partial g_{m\mu}}{\partial \vartheta^\nu} + \frac{\partial g_{m\nu}}{\partial \vartheta^\mu} - \frac{\partial g_{\mu\nu}}{\partial \vartheta^m}\right)\,. \tag{2.39}$$

Equation 2.38 can be rewritten as a second order ODE for the parameters corresponding to the geodesic path by inserting the definition of the Jacobian. However, the calculation is lengthy and not very illuminating (Do Carmo 2016). The geodesic ODE is central to the model reduction algorithm described in the next section.

2.4 Model Reduction

Model simplification through parameter reduction can be achieved for any of the global sensitivity methods described in Section 2.3.

For scanning methods such as the method of Morris (Section 2.3.1), the purpose of identifying the model parameters which have a negligible influence on the model performance metric is called “screening” or “factor fixing”. Parameters that are deemed insensitive based on the respective metric of the method can be fixed to any values of the parameter distributions used in the sensitivity analysis. For the Morris method, an arbitrary cutoff for the ℓ_2-norm of normalized Morris pairs ($\ell_2 = \sqrt{\mu^{*2} + \sigma^{*2}}$) is often used to distinguish relevant from irrelevant model parameters (Link et al. 2018). Otherwise, parameters can only be “ranked” according to their relative contribution based on the Morris mean (e.g., Pagel et al. 2014). Factor fixing will reduce the number of model parameters that have to be considered, e.g., in refined calibrations (e.g., Van Werkhoven et al. 2009), but it only makes the model conceptually simpler if the zero value is in the range of parameter variations that do not affect model output. If many parameters show interaction effects it is difficult to identify possible model reduction steps.

For Bayesian methods (Section 2.3.2) the analysis of the covariance structure of the posterior parameter distribution usually has to be complemented by an additional computational step, e.g., a representation of the likelihood-ratio test (Maiwald et al. 2016). The computation of parameter profile likelihoods in order to assess identifiability is routinely implemented in parameter estimation toolboxes, e.g., MATLAB PESTO (Stapor et al. 2017). The information on parameter limits resulting from the computation of parameter profiles has to be manually translated into algebraic model simplifications.

In contrast to the previously described methods which rely on parameter space sampling, the Manifold Boundary Approximation Method (MBAM) (Transtrum and Qiu 2014) uses a geometrical approach to model simplification involving the model manifold (Fig. 2.3). The MBAM proceeds in four computational steps:

1. At an initial parameter point on the model manifold ϑ_0, calculate the Jacobian matrix (Eq. 2.35) and construct the FIM (Eq. 2.34). The initial point is usually taken to be the best-fit parameter value, $\vartheta_0 = \vartheta^*$.

2. Compute the singular value decomposition of the Jacobian, $J_\sigma = U\Sigma V^T$, where $U \in \mathbb{R}^{M\times N}$, $V \in \mathbb{R}^{N\times N}$ and $\Sigma \in \mathbb{R}^{N\times N}$ is a diagonal matrix of singular values. The rows of V^T are the eigenvectors of $J_\sigma^T J$. Denote the eigenvector that corresponds to the smallest singular value by v_0.

3. Numerically solve the following initial value problem for the geodesic equation

$$\frac{\partial^2 \vartheta^i}{\partial \tau^2} = \sum_{j,k} \Gamma^i_{jk} \cdot \frac{\partial \vartheta^j}{\partial \tau} \frac{\partial \vartheta^k}{\partial \tau};\ \Gamma^i_{jk} = \sum_{l,m} \left(H^{-1}\right)^{il} \frac{\partial r_m}{\partial \vartheta^l} \frac{\partial^2 r_m}{\partial \vartheta^j \partial \vartheta^k}\,, \tag{2.40}$$

   using the parameter value $\vartheta = \vartheta^*$ and eigendirection $v = \partial\vartheta/\partial\tau = v_0$ as initial conditions. Identify the parameter limit ($\lim_{\tau\to\tau_b} \vartheta(\tau) = \pm\infty$) as solution to the geodesic equation (Eq. 2.40) at the manifold boundary where the FIM becomes singular.

4. Analytically evaluate the parameter limit in the model. Recalibrate the parameters of the reduced model ϑ_1 to match observations and repeat the process from step 1 with $\vartheta_0 = \vartheta_1$.

The optimal fit can usually only be inferred with large uncertainty, because parameter *equifinality* implies that many different model realizations lead to identical model behavior. This means that large regions in parameter space, which might correspond to different local minima, all map to a small confined region on the model manifold. Due to this parameter space compression (Machta et al. 2013), geodesics started from different

initial points will find the same manifold boundary. Transtrum and Qiu 2014 demonstrated that repeating the MBAM with perturbed initial parameter values but similar statistical behavior (goodness-of-fit) leads to the same reduced model, i.e. the MBAM is robust to the starting point ϑ^*.

A numerical solution to the geodesic equation (Eq. 2.40) requires the calculation of local sensitivity information (Section 2.2) and inversion of the FIM (Eq. 2.14) at every iteration step. Calculation of the Christoffel symbols Γ requires second order sensitivities. For large models it is computationally advantageous to approximate the contraction of the second derivatives of the residual vector with the geodesic velocities by a finite difference approximation of the resulting second directional derivative (Transtrum et al. 2018). Equation 2.40 then reads (Transtrum and Qiu 2016)

$$\frac{\partial \vartheta^i}{\partial \tau} = v^i \tag{2.41}$$

$$\frac{\partial v^i}{\partial \tau} = \sum_{l,m} \left(H^{-1}\right)^{il} \frac{\partial r_m}{\partial \vartheta^l} A_m(v)\,, \tag{2.42}$$

where $A_m(v)$ is the second directional derivative

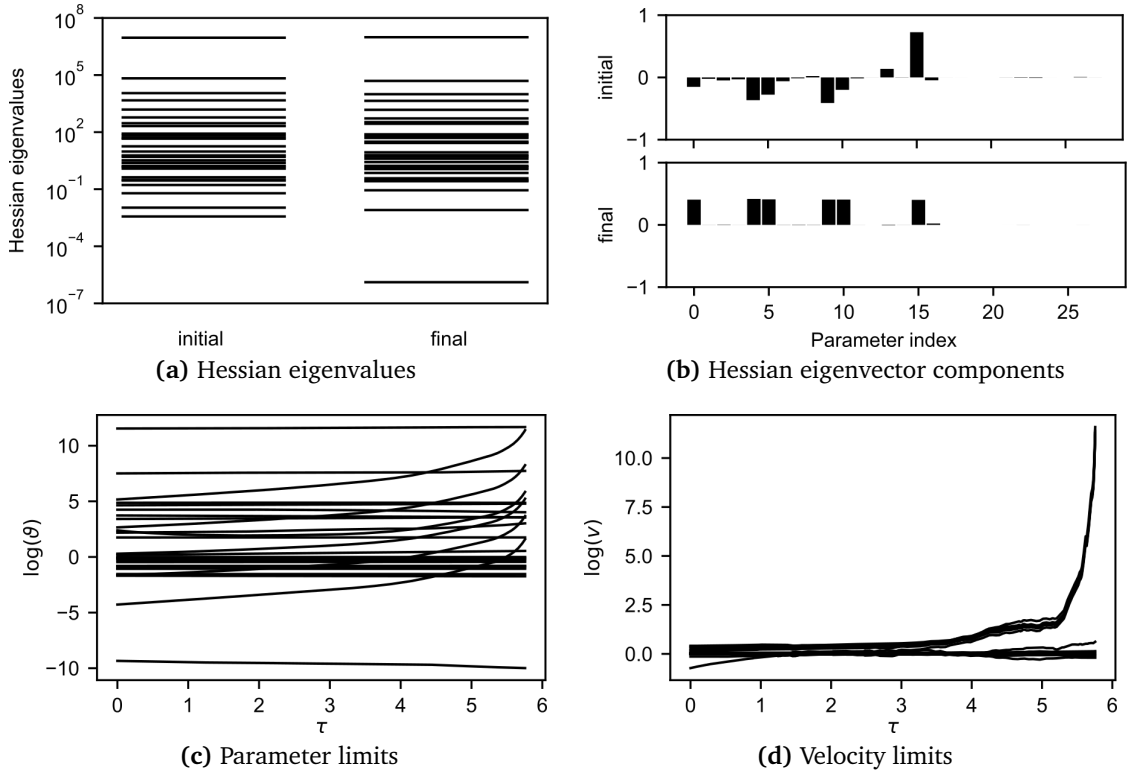
$$A_m(v) = \sum_{jk} \frac{\partial \vartheta^j}{\partial \tau} \frac{\partial \vartheta^k}{\partial \tau} \frac{\partial^2 r_m}{\partial \vartheta^j \partial \vartheta^k}\,, \tag{2.43}$$

with finite difference approximation

$$A_m(v) = \lim_{h \to 0} \frac{r_m(\vartheta + hv) + r_m(\vartheta - hv) - 2r_m(\vartheta)}{h^2}\,. \tag{2.44}$$

The geodesic ODE (Eqs. 2.41, 2.42 and 2.44) has to be integrated until a manifold boundary is identified. Since manifold widths correspond to the square root of FIM eigenvalues (Transtrum et al. 2010), the algorithm scales favorably with system size, i.e. geodesic paths to the closest boundary on the manifold are shorter for larger models. However, the eigendirection on the manifold that causes an almost imperceptible change to the model performance metric (corresponding to the smallest eigenvector v_0) can only be determined up to a constant sign from the singular value decomposition of the Jacobian matrix (either v_0 or $-v_0$). In practice, the direction is chosen in which the parameter velocity initially increases.

Four criteria can be used to discern whether a boundary has been reached (Fig. 2.4). A boundary is defined by the FIM becoming singular. As can be seen from the initial



(a) Hessian eigenvalues

(b) Hessian eigenvector components

(c) Parameter limits

(d) Velocity limits

Figure 2.4 Identifying manifold boundaries. In the semilogarithmic plots (bottom) the geodesic paths on the model manifold are parameterized by τ. For the logarithm of model parameters ϑ (c) and parameter velocities v (d), one curve is plotted per parameter (in this case 28). As the geodesic approaches a boundary (approximately at $\tau = \tau_b \approx 5.8$), six parameter values and the corresponding velocities diverge. Other model parameters slightly compensate for the limit at the boundary. The eigendirection vector (b) and FIM eigenvalues λ (a) at the start and end of the geodesic path are shown (initial/final). Once a boundary is reached, the smallest eigenvalue separates and approaches numerical zero. The final parameter space velocity vector contains only components corresponding to the parameters that take on extreme values. In this case, the geodesic ODE was integrated until the norm of the velocity vector increased by a constant factor: $|v_b| = k \cdot |v_0|$, $k = 25$.

and final plots of FIM eigenvalues, the smallest eigenvalue separates from the others and approaches numerical zero (Fig. 2.4a). The eigenvector corresponding to the smallest eigenvalue initially contains a mixture of factors, but is rotated from its initial direction to reveal the important linear combination of parameters at the boundary (Fig. 2.4b). As the geodesic approaches a boundary, model parameters asymptotically approach the limit that is defined by the boundary (in this case six parameters approach infinity, $\lim_{\tau \to \tau_b} \vartheta(\tau) = \infty$; Fig. 2.4c). Accordingly, the corresponding parameter velocities (the rates of parameter change along the manifold path) diverge (Fig. 2.4d). The increase in parameter velocity compared to the initial velocity is the most robust indicator of limiting behavior. The geodesic equation is integrated until the norm of the velocity vector has grown by a constant factor k specific to the model that is analyzed: $|v_b| = k \cdot |v_0|$.

By analytically evaluating parameter limits at the boundary, linear information is subsequently manually translated into nonlinear approximations of model behavior. Since one parameter is removed at a time, a limit involving N parameters requires the introduction of $N - 1$ new effective parameters that are expressed in terms of finite combinations of the original parameters. If parameters approach a boundary at the same rate, finite combinations correspond to cases such as "∞/∞, $0/0$, $\infty \cdot 0$, $\infty - \infty$" and might involve multiple model equations. Once the limit has been identified and analytically evaluated in the model, it is necessary to slightly recalibrate parameters in the approximate model. Denoting the reduced parameter vector by $\hat{\vartheta}$, Eq. 2.2 with $K(\hat{\vartheta})$ is used to recalibrate the approximate model with $N-1$ parameters to the data and reiterate the reduction process until all irrelevant parameters are removed from the model.

Chapter 3

Sloppiness in Soil Biogeochemical Modeling

3.1 Introduction

Soil biogeochemical modeling has progressed towards detailed, explicit representations of biophysical processes (Wieder et al. 2015), but has yet to address the associated challenges of parameter estimation (Sierra et al. 2015*a*; Wang et al. 2015; Pagel et al. 2016). The situation is not unlike in soil biology, where a universal *sloppy* structure in the parameter space of models has been reported as the underlying cause of the parameter identification problem (Waterfall et al. 2006; Gutenkunst et al. 2007; Mannakee et al. 2016). The behavior of *sloppy* models depends only on a few *stiffly* constrained nonlinear combinations of parameters. Conversely, the model behavior is insensitive to many parameter combinations that correspond to *sloppy* directions. *Sloppiness* is locally manifested in the eigenvalue spectrum of the measurement Hessian (Eq. 2.9) and globally in the effective dimensionality of the model prediction manifold (Section 2.3.3). Since soil biogeochemical models share structural elements with multiparameter models from systems biology, this chapter explores whether the reported universality of *sloppiness* (Machta et al. 2013) also extends to soil biogeochemical modeling.

3.2 Implementation

All models analyzed in this chapter were implemented in the SloppyCell software (Myers et al. 2007) in order to make use of its parameter estimation and sampling routines. Soil biogeochemical models were tested in MATLAB R2017a (MATLAB 2017) and subsequently automatically translated into SBML format for compatibility with SloppyCell (Gómez et al. 2016). For analysis of the models by Ingwersen et al. 2008; German et al. 2012; Manzoni et al. 2014; Pagel et al. 2014; Wang et al. 2015, synthetic continuous time series data for all model variables were generated that the models fit perfectly at respective published optimal parameter values ϑ^* (Tables A1, A4, A7, A10 and A13). The associated cost function is given by Eq. 2.4. Local sensitivity equations for calculating the Hessian (Eq. 2.9) and right hand side of the geodesic ODE (Eq. 2.40) were solved by the direct differential method (Section 2.2). For numerical integration of Eq. 2.1, SloppyCell uses a Python 2.7 interface to the implicit ODE solver DASKR written in Fortran (Brown et al. 1994).

For the Monod model for microbial growth (Section 3.3.3), synthetic data at four time points was generated from the "true" base parameter values $\vartheta^* = \{Y^*, V^*_{max}, K^*_S\} = \{0.5, 0.25\,1/d, 0.5\,mg\,C/g\}$. Normally-distributed noise with standard deviation equal to 10% of the respective maximum carbon concentration was added to the model output. The initial conditions were assumed to be known ($C_B(0) = 0.03\,mg\,C/g$, $C_S(0) = 1.0\,mg\,C/g$). For visualization purposes, only the maximum growth rate V_{max} and the saturation constant K_S were allowed to vary in the analysis. Log-normal priors[1] were added to the nonlinear least-squares cost function (Eq. 2.2) for the remaining two parameters in order to penalize values that stray too far from the base values ϑ^*. The Python 2.7 code of the "Robertson example" (Mannakee et al. 2016) that is distributed with the SloppyCell software (Myers et al. 2007) was modified to generate most of the results and figures in Section 3.3.3. The minimization of the nonlinear least squares cost function was performed numerically with an implementation of the standard Levenberg-Marquardt routine (Press et al. 2007). A global sensitivity analysis that relies on parameter space sampling was conducted with the SloppyCell MCMC algorithm, either with or without importance sampling (Section 2.3.2). In order to explore the model manifold (Section 2.3.3) of the Monod model, an available Python 2.7 script (Transtrum 2016) was utilized to solve the geodesic equation (Eq. 2.40). To ensure positive values and improve numerical performance, all calculations were performed on a logarithmic scale for ϑ.

3.3 Results

3.3.1 Sloppy Local Sensitivity Spectra of Soil Biogeochemical Models

The local eigenvalue spectra of five biogeochemical models evaluated at their respective published optimal parameter values are shown in Fig. 3.1a. All normalized spectra are characterized by an approximately uniform spacing of eigenvalues of the Hessian matrix over many orders of magnitude in log-space. This pattern in the local eigenvalue spectrum is the signature of a *sloppy* model (Brown and Sethna 2003).

Each eigenvalue corresponds to an eigenvector, i.e. a linear combination of parameters that determines model behavior. A few eigenvectors are oriented along *stiff* directions in parameter space in which the model is sensitive to parameter perturbations. Each successive direction in parameter space is less important for explaining model behavior by a roughly constant factor. A clear cutoff for parameter importance cannot be defined via the magnitude of the eigenvalues. Locally, the cost contours of soil biogeochemical mod-

[1] The 95% confidence interval of the normal distribution of $\log\vartheta$ with expectation value $\nu = \vartheta^*$ and standard deviation $\sigma = 1000$ is the interval $[\nu/\sigma, \nu \cdot \sigma]$.

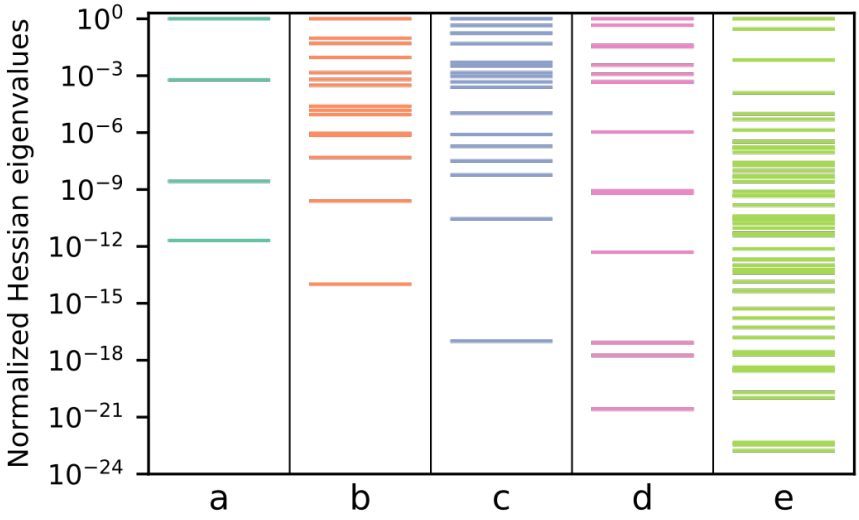
els have aspect ratios larger than one thousand. These directions in parameter space are termed *sloppy*, because associated combinations of parameters can be changed over large ranges without affecting model behavior. As a result, most individual parameter values will be difficult to infer from data. According to the Cramer-Rao bound (Eq. 2.19), inferring parameters from the *sloppiest* direction in parameter space corresponding to the smallest eigenvalue in Fig. 3.1a would require approximately 10^{22} more data than for the *stiffest* combination. This would be three times as difficult as inferring microscopic details from the diffusion equation (Machta et al. 2013).

The principal axes of the respective Hessian ellipsoids are generally not aligned with the bare parameter axes (Fig. 3.1b). This indicates that nonlinear parameter combinations compensate for each other to produce the same model output. The model behavior is not predominantly sensitive to changes in individual parameter values.

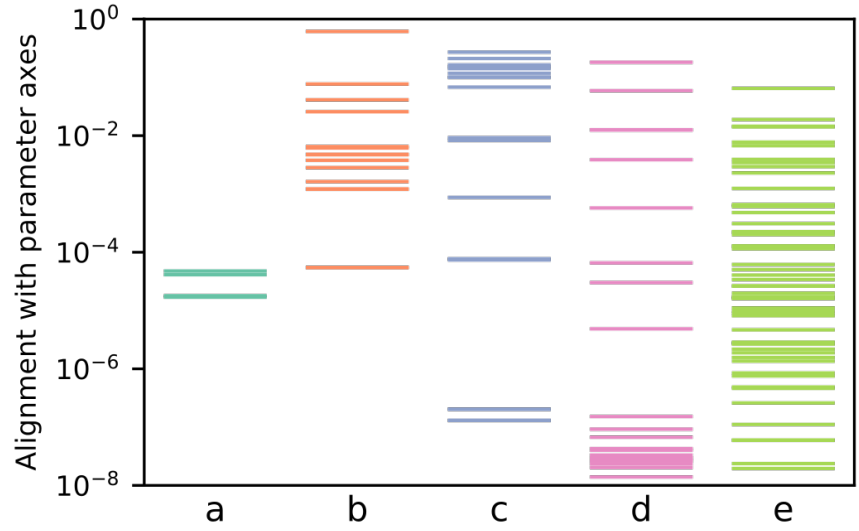
A factor analysis of the four *stiffest* eigenvectors is shown in Fig. 3.2. Except for the model by German et al. 2012, the parameters with largest factor loadings are consistent across different eigenvectors. Model reduction by the MBAM (Section 2.4) would likely identify the same limit from the first four directions in parameter space. The eigenvectors however can contain arbitrary mixtures of components from *sloppy* parameters (Gutenkunst 2007). Due to strong nonlinearities in parameter space, the factor analysis by itself is not very useful in guiding model simplification.

3.3.2 Semi-Global Analysis of the PECCAD ODE Model

Figure 3.3 paints a semi-global picture of the sensitivity of the PECCAD ODE model (Pagel et al. 2014). Column (a) depicts the Hessian eigenvalues at the best fit point (corresponding to Fig. 3.1e). Column (b) shows eigenvalue spectra generated from 10 different parameter sets drawn from an MCMC parameter ensemble. The eigenvalue spectrum is locally *sloppy* at each investigated point. Column (c) shows the average of the eigenvalues $\overline{\lambda_b}$ of the columns in (b). However, eigenvalues of distinct Hessian matrices are only additive when the corresponding eigenvectors are the same. That this is not the case can be deduced from column (d). The left spectrum in column (d) shows the eigenvalues of $\overline{H_b}$, the eigenvalues of the average of the Hessian matrices in the ensemble. The right spectrum in column (d) shows the eigenvalues of $(\overline{H_b^{-1}})^{-1}$, the inverse of the average of the inverse Hessian matrices in the ensemble. The former average conserves eigenvalues that are large in any of the individual Hessians in the ensemble. The latter conserves eigenvalues that are small. The left spectrum is condensed compared to the individual



(a) Normalized eigenvalue spectra of five biogeochemical models



(b) Alignment of eigenvectors with parameter axes

Figure 3.1 Local sensitivity analysis of soil biogeochemical models. (a) Eigenvalue spectra derived from the measurement Hessian are plotted from left to right in increasing order of apparent model complexity (the number of unknown parameters). The eigenvalues are normalized by the maximum eigenvalue. All models show a characteristic *sloppy* eigenvalue distribution at published optimal parameter values. (b) If the alignment has a value of unity it means that one principal direction in parameter space lies along a bare parameter direction, i.e. that the projection of the Hessian ellipsoid onto the corresponding axis is equal to the intersection. The spectra indicate that principal directions in parameter space are generally skewed from the bare parameter axes. Models: (a) German et al. 2012; (b) Ingwersen et al. 2008; (c) Wang et al. 2015; (d) Manzoni et al. 2014; (e) Pagel et al. 2014.

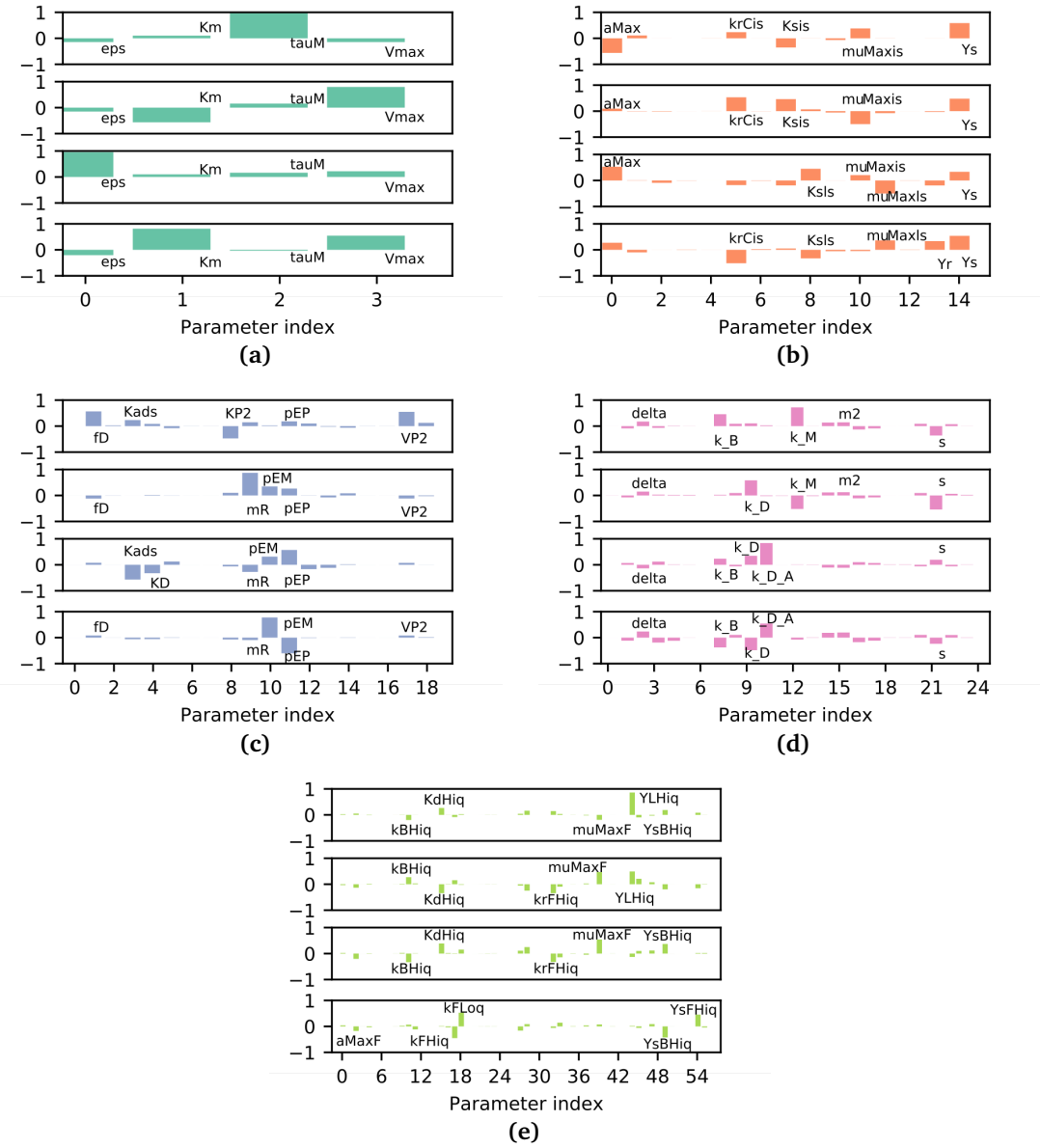


Figure 3.2 Factor analysis of the four *stiffest* eigenvalues. ((a) German et al. 2012; (b) Ingwersen et al. 2008; (c) Wang et al. 2015; (d) Manzoni et al. 2014; (e) Pagel et al. 2014). The five parameters with the largest contributions to each eigenvector are labeled.

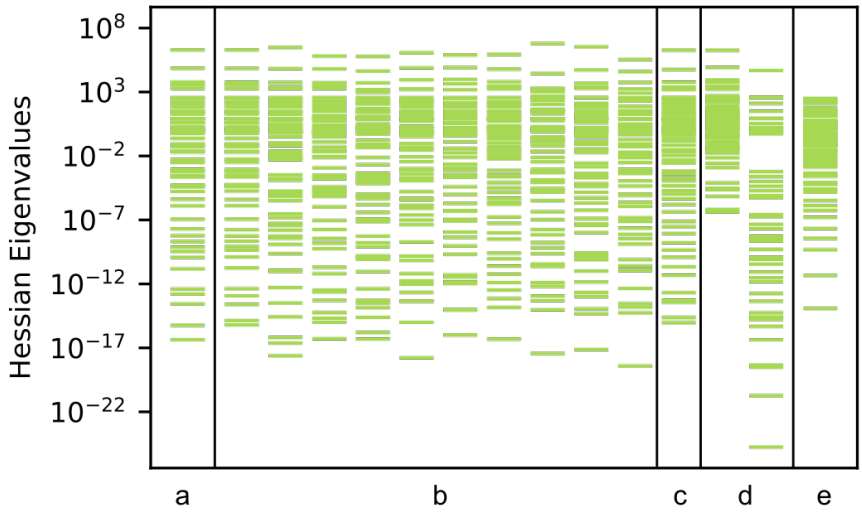


Figure 3.3 A semi-global perspective on the Hessian eigenvalues: (a) evaluated locally at the best fit point; (b) evaluated locally at ten different points close to the best fit point; (c) average $\overline{\lambda_b}$ of the eigenvalues in column (b); (d) left: eigenvalues of the average of the Hessian matrices $\overline{H_b}$ in column (b), right: eigenvalues of the inverse of the average of the inverse Hessian matrices $(\overline{H_b^{-1}})^{-1}$ in column (b); (e) eigenvalues derived from a Principal Component Analysis of the ten-point parameter ensemble in column (b).

spectra in column (b) and the right spectrum contains more small eigenvalues. Consequently, *stiff* and *sloppy* directions must change depending on the point in parameter space at which the eigenvalue spectrum is evaluated. Descriptive statistics for the eigenvalues cannot readily be computed from a set of spectra.

The eigenvalues in column (e) derive from an eigendecomposition of the sample covariance matrix of the MCMC parameter ensemble (Principal Component Analysis, PCA). By PCA, the variance of the position of points in parameter space is analyzed. In order to facilitate comparison with the eigenvalues of the Hessian, column (d) contains the inverse eigenvalues of the principal components (Eq. 2.17). When parameter nonlinearities are taken into account, the *sloppy* local picture of the cost landscape topography is confirmed by the spectrum of PCA eigenvalues.

3.3.3 Global Analysis of the Monod Model for Microbial Growth

The Monod model for microbial growth is one of the most frequently used and best studied models in microbiology (Monod 1949; Kovárová-Kovar and Egli 1998; Strigul et al. 2009). While originally developed for the growth of bacterial cultures in homogeneous batch experiments, it has become common practice to incorporate the Monod model as

a basic biokinetic component into soil biogeochemical models (Section 2.1). The Monod model describes microbial growth as a nonlinear system of ODEs with three parameters - the maximum growth rate V_{max}, the substrate saturation constant K_S and the yield coefficient Y:

$$\frac{dC_B}{dt} = \mu(t)C_B = \frac{V_{max}C_S}{K_S + C_S}C_B \tag{3.1}$$

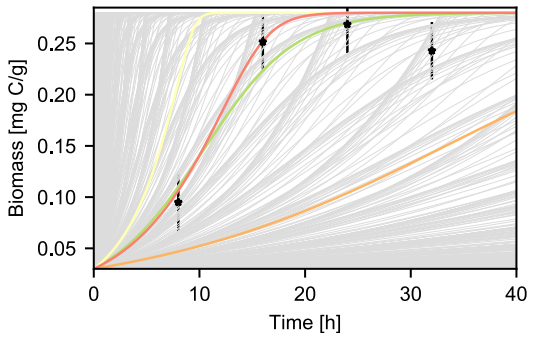
$$\frac{dC_S}{dt} = Y\mu(t)C_B = Y\frac{V_{max}C_S}{K_S + C_S}C_B\,. \tag{3.2}$$

The Monod model assumes that the microbial growth rate $\mu(t)$ depends nonlinearly on the carbon concentration C_S at time t in the system and that a constant fraction Y of the consumed carbon is transformed into microbial biomass C_B. It is equivalent to the minimal soil carbon model by German et al. 2012 described in Section 2.1, when carbon input and the first-order cycling rate of microbial biomass are neglected. In the context of soil carbon modeling, the Monod model represents a minimal *sloppy* model (Section 3.3.1).

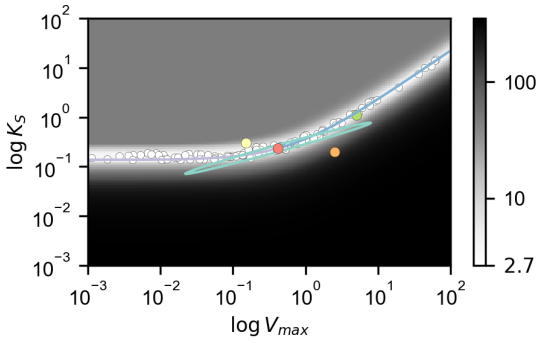
Figure 3.4a shows the model trajectory for microbial biomass C_B generated from the inferred optimal parameter values
($\vartheta^{**} = \{0.5, 0.23\,1/d, 0.42\,mg\,C/g\}, K(\vartheta^{**}) = 2.67$). Due to the noise that was added to the synthetic data, the best fit trajectory (red line) does not pass through the fourth data point, although the fit is within the error bars of the other data points. This is because model behavior is constrained by the sigmoidal form of the growth curve.

A cost surface in parameter space was generated by varying model parameters within their respective physical ranges and recording the value of the cost function (Fig. 3.4b). The topology of the cost landscape is dominated by a narrow canyon of equal goodness-of-fit. The likelihood-based confidence region appears to be infinitely extended. This indicates that parameters in the Monod model with fixed growth yield are structurally non-identifiable (compare, e.g., with Raue et al. 2009, Fig. 1A).

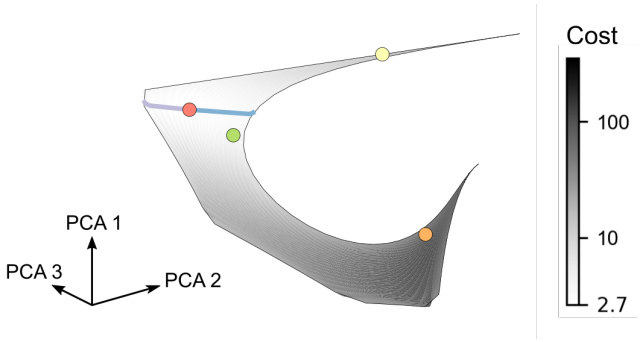
Local sensitivity is given by the eigenvalues of the measurement Hessian (Eq. 2.14) at the best fit point in parameter space ($\lambda_1 = 182.7$, $\lambda_2 = 0.4$). The ellipse of constant cost (Fig. 3.4b) illustrates two key features of *sloppiness*. First, the ellipse has a large aspect ratio, i.e. one direction in parameter space constrains model behavior much more than the other direction. Second, the ellipse is tilted from the bare parameter axes and hence constrains combinations of model parameters. The major axis of the ellipse is oriented along the eigenvector corresponding to λ_2 and shows a direction in which combinations of $\log V_{max}$ and $\log K_S$ can be changed substantially without changing the fit



(a) Time domain

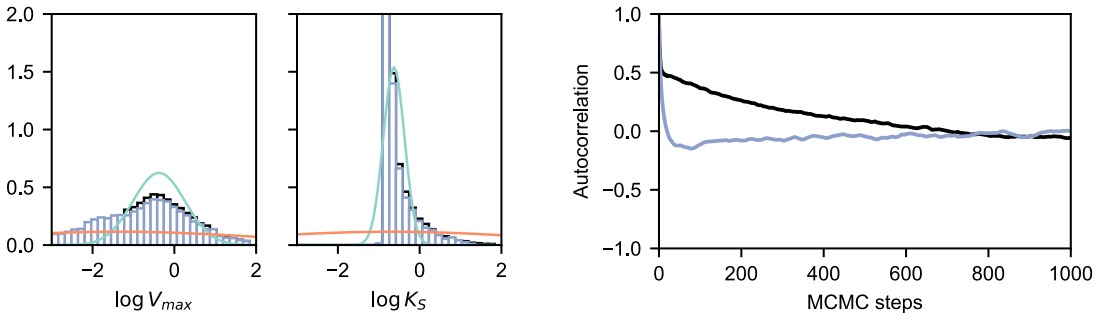


(b) Parameter space



(c) Data space

Figure 3.4 Visualization of the nonlinear least squares problem for microbial growth. Varying maximum growth rate V_{max} and substrate affinity K_S generates (a) different model trajectories, (b) a cost surface in parameter space and (c) a manifold in data space. (a) Time domain: the first three data points (black stars) and analyticity of the model constrain the predictions of the best fit trajectory (red line) at the fourth measurement point. (b) Parameter space: shown is the local approximation of the Hessian (turquoise ellipse) around the best fit point (red dot) and 120 parameter samples from an MCMC ensemble (white dots). Geodesics (purple, blue) align with the curved cost canyon in parameter space. (c) Data space: three-dimensional projection of all possible model predictions for substrate and biomass at time points for which experimental data is available. The first three principal components (PCA) explain 99.8% of the variance in the trajectory data. The geodesics connect local information at the best-fit point (red dot) with the global boundaries of the model (outlined in black). Large regions in parameter space (black) map to small regions (corners) on the model manifold. The canyon of equal cost in parameter space is mapped to a confined region on the manifold. Globally the model manifold and the local ellipse have similar aspect ratios.



(a) Marginal posterior distributions **(b)** Autocorrelations

Figure 3.5 Performance of the MCMC algorithm. Shown are results from 10,000 MCMC steps with (purple) and without (black) importance sampling run over the cost landscape shown in Fig. 3.4b. (a) Histograms of the marginal posterior distributions generated from the ensemble are compared with 95% confidence regions inferred from the Hessian approximation (turquoise line). For comparison, the prior log-normal parameter distributions are plotted in orange. (b) Autocorrelation of the cost as a function of MCMC steps. The autocorrelation time for the importance sampled ensemble is shorter by a factor of 10. Parameter samples 100 MCMC steps apart are statistically independent.

(this corresponds to a *sloppy* direction in parameter space). Conversely, the direction of the minor axis constrains optimal parameter values to lie within a small area around the best fit point (a *stiff* direction). The Monod model is *sloppy* with a highly anisotropic cost landscape, as locally indicated by the difference in magnitude of the Hessian eigenvalues.

The nonlinear relationship between model parameters is globally captured by an MCMC ensemble (Fig. 3.4b, white dots). Compared to the true posterior distribution resulting from the ensemble, the 95% confidence region derived from the quadratic approximation of the Hessian close to the best fit underestimates the variability of the values of the maximum specific growth rate V_{max} (Fig. 3.5a, left). The distribution of the substrate affinity K_S is truncated from below (Fig. 3.5a, right). The narrower marginal posterior distribution for K_S as compared to the prior distribution indicates good identifiability. A similar reduction in highest posterior density for V_{max} could not be achieved.

The curved cost canyon of the model poses computational challenges for the MCMC algorithm. The number of independent samples in a single ensemble is approximately given by the number of MCMC steps divided by the longest correlation time between members of the ensemble (Gutenkunst 2007). The correlation time of the cost was reduced by a factor of 10 by importance sampling of the MCMC sampling matrix (Fig. 3.5b).

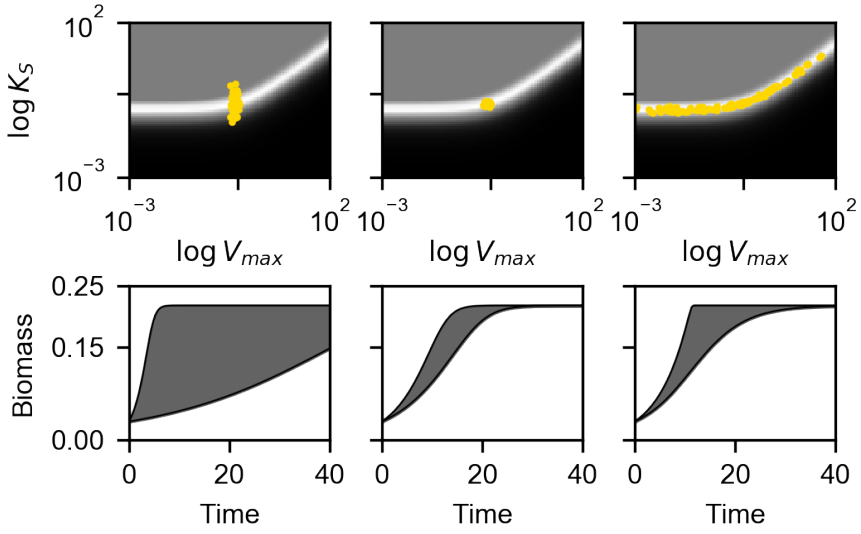


Figure 3.6 Parameter samples and associated prediction uncertainties (95% confidence intervals) for microbial growth. The figure is to be read top to bottom, left to right. Left panel: the substrate affinity K_S can only be guessed with low precision. Middle panel: maximum specific growth rate V_{max} and substrate affinity K_S can be estimated to high precision. Right panel: predictions evaluated from an MCMC parameter ensemble are informative.

The model manifold of the Monod model is embedded in an eight-dimensional data space (M=8). In order to generate a low-dimensional projection (Fig. 3.4c), model predictions over a uniform grid in parameter space (Fig. 3.4b) were evaluated. These correspond to a grid of vectors $y(\vartheta)$ in the eight-dimensional embedding space. An isometric embedding between Riemannian manifolds which preserves the metric is given by a Principal Component Analysis (PCA). The model prediction vectors were arranged in a matrix $Y = [\tilde{y}_1 \tilde{y}_2 \dots \tilde{y}_M]$, with mean shifted column vectors given by $\tilde{y}_m = y_m - \frac{1}{M}\sum_j y_j$. Given a PCA, $Y = U\Sigma V^T$, the first three principal components, $U\Sigma = YV$, explain 99.8% of the variance in the trajectory data and can be used to visualize the model prediction manifold (Fig. 3.4c). The geodesic paths and four distinct model realizations were likewise projected onto the manifold. The surface color of the manifold corresponds to the value of the nonlinear least squares cost function.

The manifold is bounded, because the cost remains finite while parameters can take on extreme values (0 or ∞). This is apparent in the cost plateaus visible in parameter space (Fig. 3.4b). As can be seen in comparison of Fig. 3.4bc, the geodesic paths can be used to explore global properties of the Monod model. The extent of the manifold in each direction coincides with the aspect ratio of the ellipse from the local quadratic

approximation. The manifold has one long axis along which model predictions change substantially and one much narrower direction. The observed empirical connection between global manifold widths and Hessian eigenvalues in *sloppy* models was formalized by Transtrum et al. 2010.

The local and Bayesian analysis demonstrate that estimating precise parameter values for K_S is difficult. How does parameter uncertainty affect model predictions? Figure 3.6 (left column) shows the impact of measurement uncertainty for K_S on model predictions for biomass when the initial substrate concentration $C_S(0)$ is reduced by 25%. If K_S can only be experimentally measured or guessed to low precision (with its 95% confidence interval spanning two orders of magnitude), large predictive uncertainty for the time course of biomass results. In order to accurately predict biomass concentrations, experimental measurements of both parameters V_{max} and K_S have to be very precise (with the 95% confidence intervals spanning only $\pm$ 50% of the true values). The required experimental precision is determined by the *stiffest* direction in parameter space (Fig. 3.6, middle column). When sampling from the MCMC ensemble (Fig. 3.6, right column), almost the same prediction uncertainty on biomass as from precise parameter measurements was recovered.

3.4 Discussion

The analysis of the simple Monod model for microbial growth illustrates many properties that are shared by more complex, *sloppy* multi-parameter soil biogeochemical models. The Monod model is *sloppy* with a highly anisotropic cost landscape and mapping between parameter and data spaces (Fig. 3.4). As reviewed below, this anisotropy has a number of important implications for soil biogeochemical modeling.

Parameter Estimation

For *sloppy* models, the inverse problem of extracting individual parameter values from fits to data by nonlinear least squares is extremely ill-posed (Gutenkunst et al. 2007). *Sloppiness* thus offers an explanation for the reported parameter identification problems in soil biogeochemical modeling (see the model descriptions in Section 2.1). The problem persists even for continuous synthetic data with very little noise that the model can fit perfectly (Fig. 3.1a). As noted by Tönsing et al. 2014, continuous time series data represent measurements that are highly correlated. As a result, densely sampled time-course data as investigated for the models in Fig. 3.1 might over-estimate real world *sloppiness*. Nevertheless, working to estimate individual parameter values in soil biogeochemical models proves difficult, as the models mostly appear to be sensitive to coordinated changes in

combinations of parameters (Fig. 3.1b).

Elaborate optimal experimental designs that allow to accurately estimate parameters of models with similar eigenvalue spectra exist in the systems biology literature, but have so far not been carried out in the lab (Apgar et al. 2010; Transtrum and Qiu 2012; Tönsing et al. 2014; White et al. 2016). The theoretical optimal experimental setups require the application of more controlled external stimuli and perturbations to the system (gene knockouts, down regulations etc.) than likely possible in the here investigated bulk soil systems. The statement with regard to model (c) in Fig. 3.1 that "more observed data points [...] would reduce parametric uncertainty" (Wang et al. 2015) is a popular misconception, since simply recording more measurements does not alleviate, but rather increase *sloppiness* (Transtrum et al. 2010).

Models with a *sloppy* eigenvalue spectrum have a low effective dimensionality (Fig. 3.4b). In higher dimensions, their model manifold is described as a *hyper-ribbon* (i.e. a geometric object that is longer than wide, wider than thick, thicker than ... by a roughly constant factor; Transtrum et al. 2011). The MBAM (Section 2.4) can be used to extract the lower dimensional effective model (i.e. reduce the model to the few long axes of the *hyper-ribbon* along which model predictions change significantly) and increase the identifiability of the remaining (effective) parameters (Transtrum and Qiu 2014; Bohner and Venkataraman 2017).

Parameter Sampling

Together, the eigenvalue spectra of different Hessian-type matrices in Fig. 3.3 indicate that *sloppiness* is a global property of the PECCAD ODE model and that the cost landscape is substantially curved. As a result, the relationship between parameters close to the best fit is nonlinear and, with increasing model complexity, likely poorly approximated by the local analysis (Fig. 3.5a). The curvature of the cost landscape poses challenges for parameter scanning algorithms.

Correlation analysis (Pianosi et al. 2016) typically relies on undirected sampling of the parameter space. Here, Latin-Hypercube Sampling (McKay et al. 1979) is the preferred method, because it avoids sampling along the bare parameter axes. For directed scanning algorithms, dealing with the "curse of dimensionality" in high-dimensional parameter spaces with complex geometries in order to study global model properties is an active problem (e.g., Zamora-Sillero et al. 2011; Leon et al. 2016). For Bayesian MCMC methods (Chib and Greenberg 1995), the Hessian matrix has to be recalculated for every member of the parameter ensemble in an importance sampling scheme (Section 2.3.2),

because *stiff* and *sloppy* directions differ between parameter samples (Fig. 3.3). Accordingly, the autocorrelation time for the cost function of the Monod model was reduced by a factor of 10 by importance sampling (Fig. 3.5b). Generating statistically independent samples in the high-dimensional, curved parameter spaces of *sloppy* soil biogeochemical models proves to be computationally expensive (Calderhead and Girolami 2011). It can be avoided by exploring global topological features of the model manifold by geodesics (Fig. 3.4c).

Model Selection

The model manifold is a global representation of a model. By exploring the model manifold with geodesics, Transtrum et al. 2010 empirically found that manifolds of *sloppy* models are bounded and exhibit a hierarchy of widths approximately given by one over the square root of the local Hessian eigenvalues. The behavior of *sloppy* models is governed by a few principal axes of the prediction manifold (Fig. 3.4c). These axes correspond to a small number n of *stiff* directions in parameter space that describe an effective model which approximates the behavior of the original model (Transtrum and Qiu 2014). The $N - n$ biogeochemical model parameters that are unconstrained by data do not contribute to model ouput variance, since the anisotropic mapping compresses large regions of parameter space into indistinguishable model predictions (small dimensions of the model manifold). The model is insensitive to coordinated changes of parameters that correspond to manifold widths smaller than one standard deviation (White et al. 2016).

In biogeochemical model selection, typically Akaike or Bayesian information criteria (IC, for an overview from a hydrological perspective, see Höge et al. 2018) are used to quantify an optimal trade-off between goodness-of-fit to data and model complexity. By construction $\mathrm{IC}(x) = C(x, \vartheta^*) + \kappa$, where $C(x, \vartheta^*)$ is the cost function (Eq. 2.2) evaluated at the maximum likelihood estimate ϑ^* and κ, the predictive complexity, is a penalty term which can be understood intuitively as the flexibility of the model in fitting data x (Choe 2017). The value of the cost function typically decreases with model dimension. In the large-sample-size limit of a structurally identifiable model, the complexity term for the AIC (BIC) is simply approximated by the number of model parameters, $\kappa = N$ ($\approx 2 \log N$), i.e. the dimensionality of the model manifold. Since the model manifold of *sloppy* models has an effective dimensionality, often much less than the number of model parameters, the AIC (and BIC) approximations for model complexity systematically overestimate the true predictive complexity at finite sample size (LaMont and Wiggins 2015, Section 5.3). The $N - n$ model parameters that are unconstrained by data do not contribute to model flexibility.

Likewise, Jeffrey's prior is not an uninformative Bayesian prior choice for *sloppy* models (Mattingly et al. 2018). Since Jeffreys prior is proportional to the volume element of the model manifold as measured by the FIM, $\sqrt{|g|}$ (Section 2.3.3), its weight along the relevant directions always depends marginally on its remaining support on the irrelevant *sloppy* dimensions of the manifold. Hence it depends on parameters to which the model is almost imperceptibly sensitive and which could be introduced arbitrarily. Instead, the prior for which the data brings the maximal amount of information is discrete, with support only on the boundaries of the model manifold (Mattingly et al. 2018, Fig. 4). It thus selects precisely the effective model that can systematically be constructed by the MBAM.

Model Prediction

Predictions from ensemble fits in *sloppy* models can be tightly constrained, despite large uncertainties in individual parameter values
(Fig. 3.6, right column). The reverse statement is that tight parameter estimates constrain predictions just as well as ensemble fits (Fig. 3.6, middle column). For model uncertainty analysis, it is hence not necessary to insist upon tight prior estimates for all model parameters, which are difficult to obtain in soil biogeochemical systems (e.g., Wang et al. 2012, 2013, 2015). The information provided by constraining n relevant parameters in *sloppy* models is enough to make informative predictions.

In the context of systems biology, it was shown that it is easier to design optimal experiments that constrain model predictions than individual parameters (Casey et al. 2007). Similar arguments have recently shifted the focus from the identification of model parameters to model predictions in the hydrological modeling community (Hermans 2017). As pointed out by Hermans 2017, it is imperative that the data is informative regarding the predictions of interest, i.e. that data and model predictions are driven by the same model parameters. For *sloppy* models, this is the case if the effective model constructed by the MBAM contains the relevant mechanisms that drive the forecast. If the relevant mechanisms are contained in the $(N - n)$ - dimensional *sloppy* subset, the model structure has to be revised. In this way, systematic errors in model structure were discovered, e.g., in density functional theory (Mortensen et al. 2005; Nikšić and Vretenar 2016; Nikšić et al. 2017). In Chapter 5, the MBAM is used to assess the representation of microbial processes associated with pesticide degradation in the PECCAD ODE model.

Most of the implications of *sloppiness* related to parameter estimation, parameter sampling, model selection and model prediction exist in some form or another in the systems biology or hydrology literature (e.g., Engl et al. 2009; Zamora-Sillero et al. 2011; Schöniger et al. 2014; Hermans 2017). Soil biogeochemical model development, how-

ever, has been struggling to implement theoretical insights from these fields (Wieder et al. 2015). The *sloppy* systems approach provides a novel perspective on problems associated with biogeochemical modeling.

Chapter 4

Problems with Parameter Inference from Heterotrophic Respiration Data

4.1 Introduction

Kinetic respiration analysis is widely used to independently determine estimates for subsets of biophysiological parameters of soil biogeochemical models (Anderson and Domsch 1978; Colores et al. 1996; Blagodatsky et al. 2000; Wutzler et al. 2012; Blagodatskaya and Kuzyakov 2013; Wang et al. 2014*a*). In a kinetic respiration assay, a soil sample is amended with carbon growth substrates (e.g., glucose and mineral nutrients) and the increase in soil heterotrophic respiration rate is monitored over time. When carbon substrates are initially in excess, microbial growth is unlimited and an exponential increase of the respiration rate is observed. The question addressed in this chapter is whether biophysiological parameters related to growth and activity of microorganisms can be reliably inferred from the exponential respiration curve by nonlinear least squares regression.

The minimal set of parameters to be inferred from respiration assays found in the literature consists of the initial microbial biomass, its active fraction and the maximum specific growth rate. The results of parameter estimation from existing studies are not directly comparable, because the approach depends on the underlying microbial physiology model. In this chapter, parameter inference from heterotrophic respiration data for the microbial physiology compartment of the MEND model (Wang et al. 2014*a*, 2015; Section 2.1) is analyzed.

The MEND microbial physiology model (Wang et al. 2014*a*) is expressed as a system of three ODEs that describe the rate of change of carbon substrate S, microbial biomass B and active microbial biomass fraction $r = B_A/B$, where B_A denotes active microbial biomass:

$$\frac{dS}{dt} = -\frac{1}{Y_G} \cdot \frac{\phi(S)}{\alpha} \cdot m_R \cdot r \cdot B \tag{4.1}$$

$$\frac{dB}{dt} = \left(\frac{\phi(S)}{\alpha} - 1 \right) \cdot m_R \cdot r \cdot B - (\beta m_R) \cdot (1 - r) \cdot B \tag{4.2}$$

$$\frac{dr}{dt} = m_r \cdot \left[(\phi(S) - r) + \left(\frac{\phi(S)}{\alpha} + \beta - 1 \right) \cdot r \cdot (1 - r) \right] . \tag{4.3}$$

Substrate uptake by microbial biomass depends on the substrate saturation level $\phi(S) = S/(K_S + S)$, where the parameter K_S is interpreted as the half-saturation constant of Monod kinetics (Section 3.3.3) and Y_G is the true growth yield. The growth and maintenance functions of active biomass are characterized by m_R, the specific maintenance rate, and $\alpha = m_R/(\mu_G + m_R)$, the ratio of m_R to the sum of maximum specific growth rate μ_G and maintenance rate m_R. The transition from active to dormant state of microbial biomass in the MEND model is determined by maintenance requirements. The fraction of dormant biomass $(1-r)\cdot B$ is not capable of growth on carbon substrate concentration S, but maintains cellular function in a dormant state at a lower maintenance cost of $\beta \cdot m_R$, where $\beta < 1$.

The fraction of carbon substrates not incorporated into active microbial biomass is respired as CO_2 and constitutes the heterotrophic respiration flux rate:

$$\nu(t) = \frac{dCO_2}{dt} = \frac{1 - Y_G}{Y_G} \cdot \frac{\phi(S)}{\alpha} \cdot m_R \cdot r \cdot B\,. \tag{4.4}$$

When substrate is in excess, $S \gg K_S$, the substrate saturation level plateaus, $\phi(S) \to 1$, and $\nu(t)$ can be expressed as an explicit function of time (Wang et al. 2014*a*, Eq. 14d):

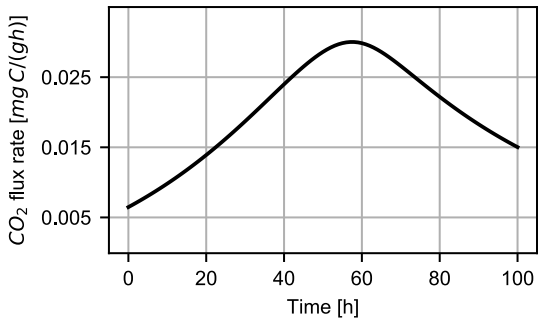
$$\nu(t) = \frac{B_0(1 - Y_G)}{Y_G} \cdot \left([\mu_G \cdot r_0 + m_R] \cdot e^{\mu_G \cdot t} - [m_R \cdot (1 - r_0)] \cdot e^{-m_R \cdot t}\right)\,. \tag{4.5}$$

The heterotrophic respiration rate is determined by the difference of two exponential functions, $e^{\mu_G \cdot t} = e^{(1/\alpha - 1)\cdot m_R \cdot t}$ and $e^{-m_R \cdot t}$. The
exponentially-increasing initial phase of heterotrophic respiration rate data from substrate-induced respiration assays (Fig. 4.1) can in principle be used to estimate five biophysiological parameters from the model: the initial total microbial biomass B_0, the initial fraction of active biomass r_0, the true growth yield Y_G, the specific maintenance rate m_R and the maximum specific growth rate μ_G via its dependence on α.

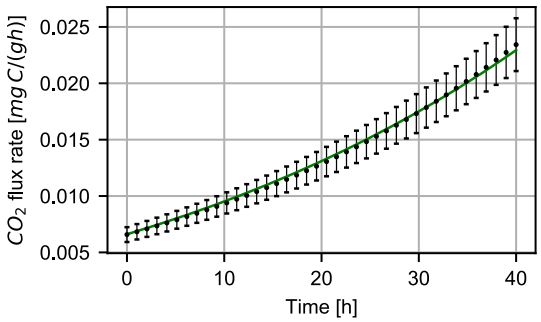
4.2 Implementation

The code used to generate the synthetic data, fit the model and perform the MBAM model reduction for the respiration assay was adapted from `https://github.com/gbohner/MBAM` (Bohner and Venkataraman 2017) and updated to be compatible with the Julia v0.6.4 release (Bezanson et al. 2017). The automated code to reproduce all figures in this chapter is available from `https://github.com/giannamars/MBAM-RESP`.

The model defined as $y = f(\vartheta, t)$, explicitly given by (Eq. 4.5), with single output $y \in \mathbb{R}$



(a)



(b)

Parameter	Unit	Base value	Fit value
B_0	$[mg\,C/g]$	0.504	0.477
r_0	$[-]$	0.394	0.052
Y_G	$[-]$	0.50	0.09
μ_G	$[-]$	0.027	0.027
m_R	$[1/h]$	0.00613	0.00038

(c)

Figure 4.1 Synthetic heterotrophic respiration data generated from the MEND microbial physiology model. (a) Exponentially-increasing and non-exponentially-increasing respiration rate phases simulated over 100 hours. Simulations were generated from the full model (Eq. 4.4) with base parameter values as in (c). (b) Exponentially-increasing respiration data sampled every hour over the first 40 hours from simulations of the full model in (a). The error bars represent 10% deviation from the noiseless synthetic data. The data is extremely well fit by the exponential model (Eq. 4.5, green solid line), but optimal parameter values from a randomly chosen optimization run deviate substantially from the "true" base values for three out of the five biophysiological parameters (fit values in (c)).

and model parameters $\vartheta = \{B_0, r_0, Y_G, \mu_G, m_R\} \in \mathbb{R}^N$, $N = 5$, was simulated using the base parameter values ϑ^* given in Fig. 4.1c. The parameter values were chosen to match the original study of the MEND physiology model by Wang et al. 2014*a*, Table 1. Noisy data was generated by multiplying the base parameter simulation with a stochastic term. The stochastic values were drawn from the distribution *Uniform*$([1-\sigma, 1+\sigma])$, $\sigma = 0.1$. The Levenberg-Marquardt solver (Press et al. 2007) was used to minimize the nonlinear least squares cost function (Eq. 2.2). A local multi-start optimization method was chosen in this study, because it performs well in terms of finding the global parameter minimum in biological systems as compared to particle swarm (or hybrid) optimization algorithms used in the original analysis of the respiration assay (Raue et al. 2013; Stapor et al. 2017, Supplementary Fig. S2). Each minimization was performed from 24 perturbed initial parameter vectors from which the global best-fit parameters were selected. Bounds on the relative size of the 95% confidence intervals of individual model parameters ϑ_i (Eq. 2.18, Gutenkunst et al. 2007; Apgar et al. 2010) were calculated from the trace of the inverse Hessian matrix H as

$$\Sigma_i = \exp\left(4 \cdot \left(\frac{\sigma^2}{T} H_{ii}^{-1}\right)^{\frac{1}{2}}\right) - 1\,, \tag{4.6}$$

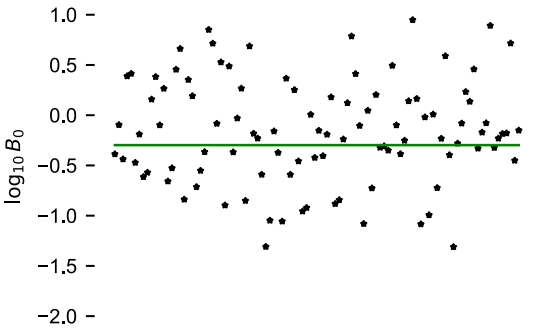
with time grid size $T = 40$. Σ_i gives the ratio of values for parameter i at the upper and lower bound of the 95% confidence interval. The coefficient of variation (CV) of a parameter distribution, defined as the ratio of the standard deviation to the mean, was likewise reported as an uncertainty measure.

Conceptual details of the model reduction by the MBAM can be found in Section 2.4. Numerical details are made available through inspection of the online simulation code. Full algebraic details of the model reduction steps are presented in Section 4.3.

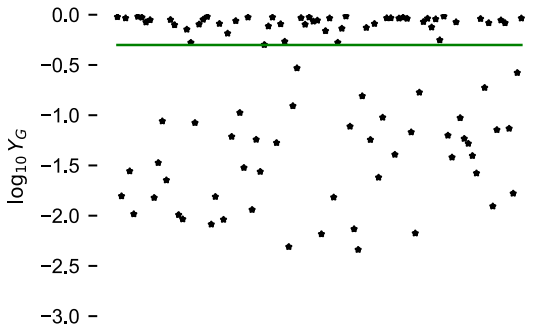
4.3 Results

4.3.1 An Illustration of Parameter Non-Identifiability

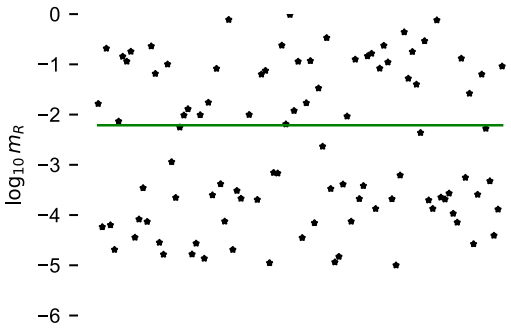
Noisy synthetic data generated from the MEND physiology model
(Eq. 4.5) are well fit (within the 10% error bars) by one instance of parameter values that deviate from the base values used to generate the data (Fig. 4.1bc). In order to determine the variation in inferred parameter values, the model was fit with a multi-start gradient-descent algorithm to 100 noisy synthetic data sets (Fig. 4.2). Except for the maximum specific growth rate, the inferred best-fit parameters span many orders of magnitude (Fig. 4.2a-e). Parameter identifiability was quantified by computing bounds on the 95% confidence interval (relative error) and by determining the ratio of the biased standard deviation to the mean of each distribution (coefficient of variation, CV). Except for the



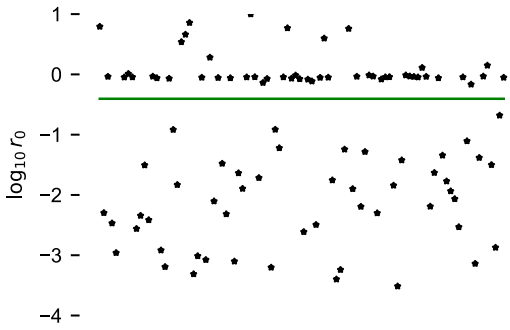
(a)



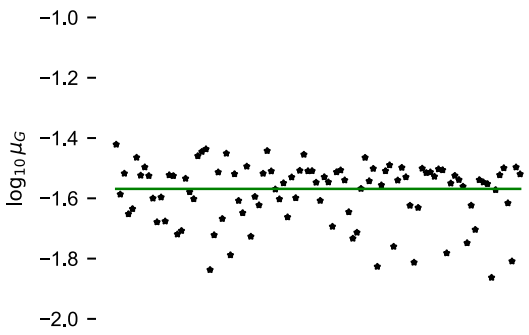
(b)



(c)



(d)



(e)

Parameter	Relative error [%]	CV [%]
B_0	Inf	3.1
r_0	Inf	1.2
Y_G	Inf	1.0
μ_G	$1.31e^{48}$	0.06
m_R	Inf	0.64

(f)

Figure 4.2 An illustration of non-identifiability. (a)-(e) Logarithm of fitted parameter values to 100 noisy synthetic data sets generated from the "true" base parameter values (green line). The values span many orders of magnitude. (f) The lower bounds on the 95% confidence intervals (relative parameter errors) are infinite. The coefficient of variation (CV) is smallest for the maximum specific growth rate μ_G.

maximum specific growth rate, the confidence intervals of parameters are infinite. Their values can only be determined with high uncertainty as confirmed by the computed CVs (Fig. 4.2f). Relative to the base parameters, the values for the true growth yield and the initial fraction of active biomass were systematically overestimated (Fig. 4.2bd). Overall, the full MEND physiology model suffers from non-identifiability.

4.3.2 Model Parameters are Non-Identifiable Because of Sloppiness

The bias in the inferred values for the true growth yield and the initial fraction of active biomass suggests that non-identifiability arises because parameter values compensate for each other to produce similar output, rather than because model output is insensitive to variations of individual model parameters. The same characteristic has been observed for *sloppy* models that locally exhibit an exponential range of parameter sensitivities (Gutenkunst et al. 2007; Machta et al. 2013; Chapter 3) and models that suffer from (more loosely-defined) parameter *equifinality* issues (Beven 2006). *Sloppiness* is defined via the eigenvalues of the Hessian matrix. Figure 4.3a shows that the square root of eigenvalues of the MEND physiology model are exponentially spaced, corresponding to a linear spacing in the logarithm. Inferring parameter combinations corresponding to successive eigenvalues in the spectrum is more difficult by a constant factor. The same trend in the spectrum is observed when the respiration data generated from Eq. 4.5 is log-transformed. Linear regression can then be used to infer parameter values. However, the log-transformed model exhibits the same degree of *sloppiness*. It hence possesses the same number of identifiable parameters and there is no information gain in the transformation.

4.3.3 Model Reduction Results in Identifiable Parameters

The MBAM was used to construct a reduced MEND physiology model the parameters of which are all identifiable and still confer mechanistic information about the system. The algorithm proceeds by removing the smallest eigenvalue of the *sloppy* spectrum in each iteration (Fig. 4.3b), whilst keeping the goodness-of-fit of the reduced models to respiration data constant (Fig. 4.5a). The output of MBAM iterations 1-3 are shown in Fig. 4.4 (left column). MBAM calculates a trajectory through parameter space, parameterized by τ, for which the model fit remains constant. Conversely, each "time point" τ corresponds to a set of parameter values that give rise to equivalent model behavior. In the extreme case, parameter values in log-space diverge, i.e. parameters take on values close to 0 or ∞. Once divergent parameters are identified, the algorithm is terminated and the model is reparameterized in terms of the emergent finite parameter combinations. Through reparameterization of the model, one original model parameter is removed per iteration.

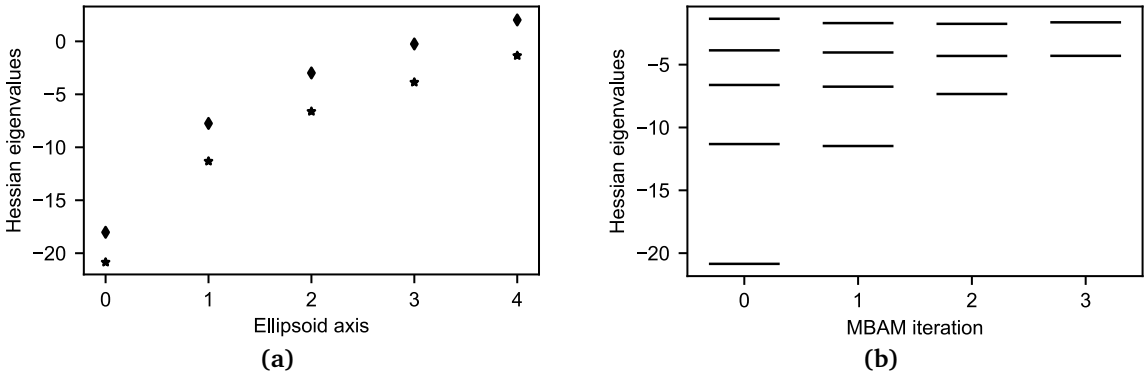


Figure 4.3 The MEND physiology model is *sloppy*. (a) Calculated logarithm of the square root of Hessian eigenvalues for the original (stars) and log-transformed (diamond) data. The linear trend in log-space between Hessian eigenvalues and the number of ellipsoid axes is the signature of a *sloppy* model. (b) The MBAM iterations remove the parameter combination corresponding to the smallest eigenvalue from the model.

For the MEND physiology model, this procedure was repeated four times.

While the MBAM algorithm is deterministic and fully automated for the example in this chapter, the 0th model reduction step was easily guessed. Recall that the full model is given by

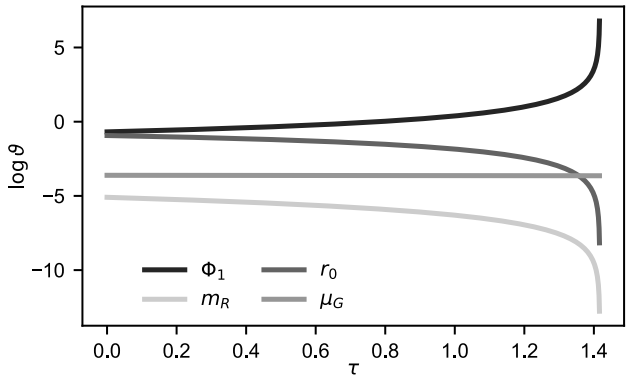
$$\nu(t) = \frac{B_0(1-Y_G)}{Y_G} \cdot \left([\mu_G \cdot r_0 + m_R] \cdot e^{\mu_G \cdot t} - [m_R \cdot (1-r_0)] \cdot e^{-m_R \cdot t} \right) . \tag{4.7}$$

The prefactor of the difference of exponentials is composed entirely of parameters. When B_0 and Y_G are of the same order of magnitude, they can be varied without changing the effective value of the prefactor. In fact, the 0th limit corresponds to $B_0, Y_G \to 0$ as $\tau \to \infty$ (not shown). Reparameterizing the model by introducing the effective parameter $\Phi_1 = B_0/Y_G$ removes the smallest Hessian eigenvector from the spectrum (Fig. 4.3b) without changing the fit to data (Fig. 4.5a). After the 0th iteration, the model reads

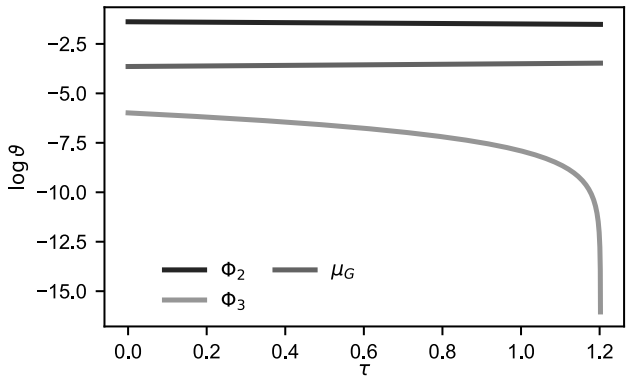
$$\nu_1(t) = \Phi_1 \cdot \left([\mu_G \cdot r_0 + m_R] \cdot e^{\mu_G \cdot t} - [m_R \cdot (1-r_0)] \cdot e^{-m_R \cdot t} \right) . \tag{4.8}$$

The first MBAM iteration subsequently sends $\Phi_1 \to \infty$ and $r_0, m_R \to 0$ as $\tau \to \tau_b = 1.4$ (Fig. 4.4, first row). After the first reduction the model is given by

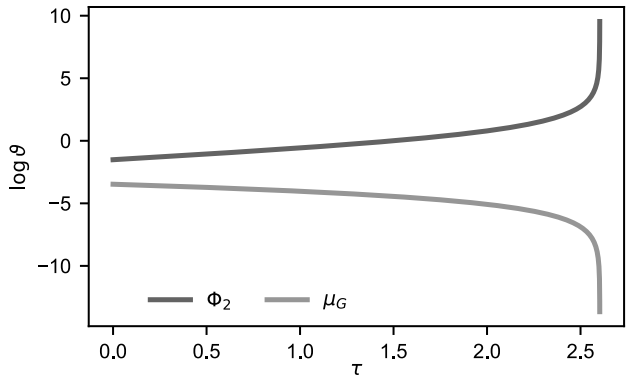
$$\nu_2(t) = [\Phi_2 \mu_G + \Phi_3] \cdot e^{\mu_G \cdot t} - \Phi_3 \, . \tag{4.9}$$



Parameter	Relative error [%]
Φ_1	Inf
Y_G	Inf
μ_G	$1.31e^{48}$
m_R	Inf



Parameter	Relative error [%]
Φ_2	$2.47e^{7}$
Φ_3	$2.44e^{43}$
μ_G	$3.26e^{8}$



Parame-ter	Relative error [%]	CV [%]
Φ_2	6111.8	0.03
μ_G	786.4	0.008

Figure 4.4 Intermediate MBAM steps 1-3 (left column) and relative parameter errors (right column). The left column shows parameter limits in each MBAM step that are identified at the end of the geodesic path parameterized by τ. The corresponding table in the same row shows the relative parameter errors derived from the inverse Hessian approximation at the start of the reduction. The MBAM removes the least-identifiable parameter combination.

Finite *emergent* parameter combinations correspond to $\Phi_2 = \Phi_1 \cdot r_0$ and $\Phi_3 = \Phi_1 \cdot m_R$, i.e. parameter combinations that can be combined to the product "$0 \cdot \infty$".

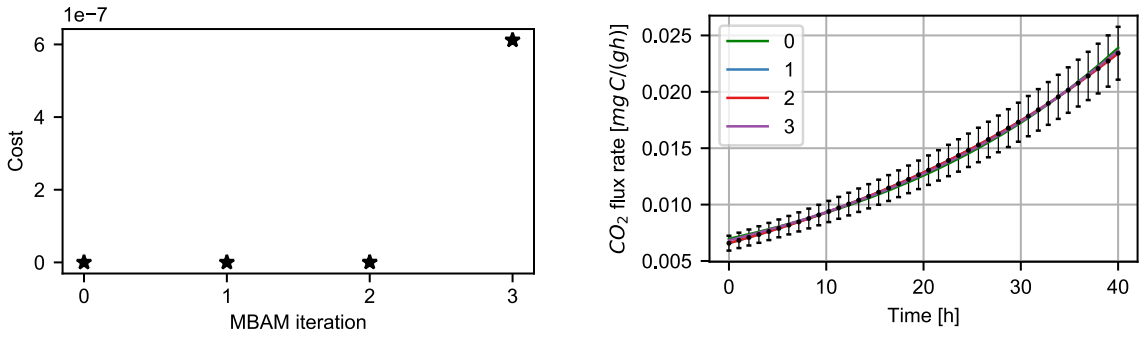
The second limit sends $\Phi_3 \to 0$ (Fig. 4.4, second row) and completely removes the decaying exponential term from the MEND physiology model:

$$\nu_3(t) = [\Phi_2 \mu_G] \cdot e^{\mu_G \cdot t} \,. \tag{4.10}$$

Notice that in the second iteration the other two parameters Φ_2 and μ_G have to slightly compensate for the removal of Φ_3, while the first reduction took place without significant compensation by μ_G.

The last iteration sends $\Phi_2 \to \infty$ and $\mu_G \to 0$ (Fig. 4.4, third row), but results in a constant heterotrophic respiration rate, $\nu_4 = \Phi_5 = \Phi_2 \mu_G = \text{const}$, which is clearly not capable of matching the data.

All intermediate reduced models, $\nu_i(t),\ i \in \{1, 2, 3\}$, fit the data extremely well (Fig. 4.5). Lower bounds on the relative parameter errors calculated at the start of MBAM consistently decrease with each iteration (Fig. 4.4, right column). Each model reduction leads to a smaller number of non-identifiable parameter values, while the identifiability of single parameters (e.g., Φ_2 and μ_G) increases. Accordingly, the value of the maximum specific growth rate base parameter μ_G as inferred by the reduced model $\nu_3(t)$ from 100 noisy data sets varied one order of magnitude less than when inferred from the full model



(a) **(b)**

Figure 4.5 Goodness-of-fit to heterotrophic respiration data (a) and visualization of the fitted respiration curve (b) for each reduced model $\nu_i(t), i \in [0, 3]$ (Eqs. 4.7 to 4.10). The increase in the cost function value for the third reduction is negligible. All models fit the data extremely well.

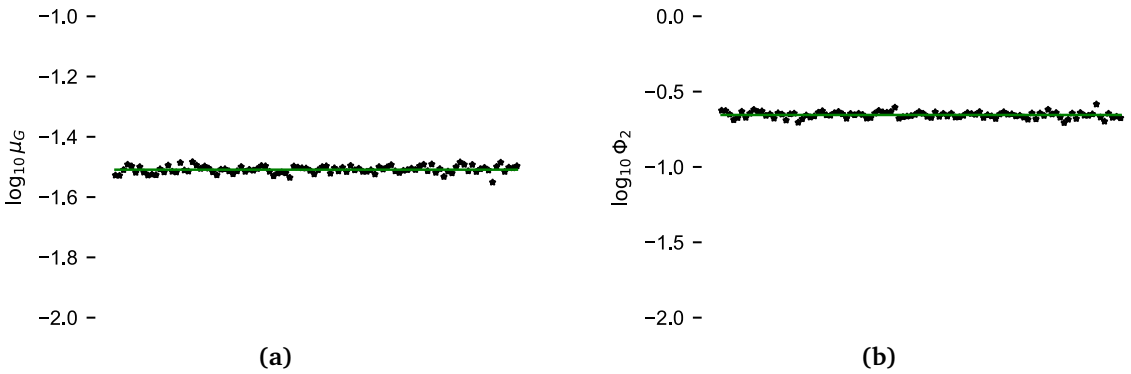


Figure 4.6 An illustration of identifiability. Logarithm of fitted parameter values of the reduced model ν_3 (Eq. 4.10) to 100 noisy synthetic data sets generated from the "true" base parameter values (green line). The coefficient of variation for each parameter is smaller as compared to the full model (Fig. 4.2).

(compare Fig. 4.2e and Fig. 4.6a, as well as the coefficients of variation in Fig. 4.2f and Fig. 4.4 (bottom row)). Differences in base and inferred parameters of the full model that give rise to the same model output keep the *emergent* parameter $\Phi_2 = B_0 r_0 / Y_G$ constant. This is the explanation for the observed bias in the estimation of the base values of parameters r_0 and Y_G from the full model (Fig. 4.2bd). Conversely, only the *emergent* parameter combination Φ_2 can be reliably inferred from the data.

4.4 Discussion

Models of the form $y(t) = Ae^{\vartheta_1 t}$ and $y(t) = A(e^{\vartheta_1 t} + e^{-\vartheta_2 t})$ are structurally identifiable for generic positive parameter values (except for a pathology at $A = 0$). Structural non-identifiability means that model parameters cannot be identified from an infinite amount of noiseless data, i.e. parameters can never be estimated (Villaverde et al. 2016). Structural identifiability is a prerequisite for practical identifiability studied in this chapter. Practical identifiability refers to quantifying the uncertainty in parameter values when estimated from noisy data. The nonlinear least squares problem of fitting sums of exponentials to data has been studied extensively at least since Bates and Watts 1988; van den Bos and Swarte 1993. The problem is notoriously ill-posed. In fact, it is the prototype of a sloppy model (Transtrum et al. 2010, 2011). E.g., Mattingly et al. 2018, Fig. 4 show that the difference between one or two decay rate parameters ϑ_i being identifiable from the sum of two exponentials can be the difference between a week's data and a year's data as determined by the ratio of the data noise to the number of measurement repetitions.

In the context of classical soil carbon modeling, the problem has recently been associated with kinetic fitting of incubation data (Weihermüller et al. 2018). Similar to the heterotrophic respiration assay analyzed in this study, soil properties from classical incubation experiments have to be inferred from exponentially increasing cumulated CO_2 fluxes. E.g., for a double-pool model, the total organic carbon pool is separated into two different compartments (a labile pool with short turnover time and a recalcitrant pool with longer turnover time). The two exponential decay constants ϑ_i correspond to kinetic parameters of carbon turnover pools that can directly be incorporated into complex, predictive, multiple timescale carbon models (Jenkinson 1990; Parton et al. 1987; Wutzler and Reichstein 2013). While the authors proposed statistical guidelines for the estimation of kinetic parameters, they concluded that "all these recommendations will not overcome the ill-posed problem, which is inherent to the experimental data and the nature of the simple carbon decay (double-pool) model" (Weihermüller et al. 2018, p.269). Similarly, Wutzler et al. 2012 focused on how the uncertainty of biophysiological parameter estimation from kinetic respiration analysis is influenced by measurement uncertainty and data selection. Due to the degree of difficulty of the inverse problem however, the uninformed investigation of available respiration data is unlikely to lead to universal guidelines for parameter identification in the nonlinear least squares problem.

In contrast to questions of data selection, this study focused on overcoming the ill-posed nature of the inference problem by constructing reduced models with identifiable parameters. In line with the original study (Wang et al. 2014*a*), the MBAM identifies the maximum specific growth rate μ_G as a model parameter that can reliably be estimated from the heterotrophic respiration assay. Model reduction decreases the uncertainty of the inferred parameter value as evidenced by a significant decrease in the coefficient of variation for the maximum specific growth rate in each reduction step.

In contrast to the original study, the MBAM iteration shows that the exponentially-increasing respiration curve cannot be used to reliably determine the initial active microbial biomass. Instead, the second identifiable parameter $\Phi_2 = B_0 r_0 / Y_G$ is a nonlinear combination of the initial total biomass B_0, the initial active microbial fraction r_0 and the true growth yield Y_G. Hence, only the ratio of the initial active biomass to the true growth yield can be inferred with high certainty. This result is non-trivial and cannot be discerned from the individual relative parameter errors or reported coefficients of variations of the full MEND physiology model (Fig. 4.2f).

Since the *emergent* parameter combination contains an initial condition of the MEND model, it does not convey mechanistic information about the system under study, as, e.g.,

observed in the context of other biological systems (Transtrum and Qiu 2016; Bohner and Venkataraman 2017). In particular, a trade-off between initial active biomass and true growth yield is neither biophysically possible, nor biophysiologically reported (Martiny et al. 2015; Litchman et al. 2015). In this case, the *emergence* is spurious, rather than a meaningful result of the nonlinear least squares problem. The *emergent* parameter cannot be directly integrated into more complex soil carbon models.

Reporting the uncertainties of estimated parameters is as important as reporting the best estimates for predictive soil carbon modeling (Raupach et al. 2005; Wieder et al. 2015; Luo et al. 2016). In order to increase confidence in selective parameter estimates and associated uncertainties, the MBAM could easily be used as an add-on to the model fitting process in kinetic respiration assays. A holistic way forward in the analysis of heterotrophic respiration data could be to revert back to the noisy respiration data sets used in the original studies (Wutzler et al. 2012; Weihermüller et al. 2018) and contrast results of regularization methods for nonlinear least squares problems (Engl et al. 1996, 2009; Gábor and Banga 2015) on the identifiability of biophysiological parameters with results of the MBAM model reduction.

Chapter 5

Emergent Controls of Pesticide Degradation

5.1 Introduction

As the most complex representative among the biogeochemical models, the PECCAD model (PEsticide degradation Coupled to CArbon turnover in the Detritusphere; Pagel et al. 2014) that simulates degradation of the herbicide 4-chloro-2-methylphenoxyacetic acid (MCPA) coupled to carbon (C) turnover in soil is analyzed in this chapter. The model was originally formulated with the aim to identify regulation mechanisms of accelerated pesticide degradation in soil in response to supply of fresh C from decomposing plant litter. As shown in Fig. 5.1 the model can be observed at different data resolutions.

PECCAD couples the dynamics of two pesticide pools (dissolved C_P and sorbed phase C_{P-s}) to that of several C pools (readily available high quality C C_{hiq}, and sorbed phase C_{hiq-s}, recalcitrant low quality C C_{loq} and sorbed phase C_{loq-s}, insoluble soil organic matter C_I) and microbial populations (bacteria C_B, fungi C_F, specific pesticide degraders C_{BP}). The model simulates the physiological state of microorganisms ($r_i, i \in \{B, F, BP\}$) to account for active and dormant biomass. Input of litter-derived dissolved organic C (DOC) and partitioning into high and low quality fractions is simulated with a time-dependent empirical litter decomposition function (Pagel et al. 2014, Online Resource 3). Model parameters of PECCAD that have to be inferred from data can be loosely grouped into biokinetic parameters (maximum growth and decomposition rates, substrate affinity coefficients, substrate uptake efficiencies), physicochemical parameters (sorption coefficients, partitioning coefficients of C pools) and unknown initial conditions (physiological state index variables).

5.2 Implementation and Data

In contrast to the original PECCAD model, spatial variations of state variables and transport processes were neglected in this study. That is, the original system of 12 coupled partial and ordinary differential equations (Pagel et al. 2014, Table 1-2) was transformed to an ODE system (PECCAD ODE). This was needed to stay within the computational framework of the other analyzed biogeochemical models.

PECCAD ODE was implemented in the SloppyCell software (Myers et al. 2007). Equa-

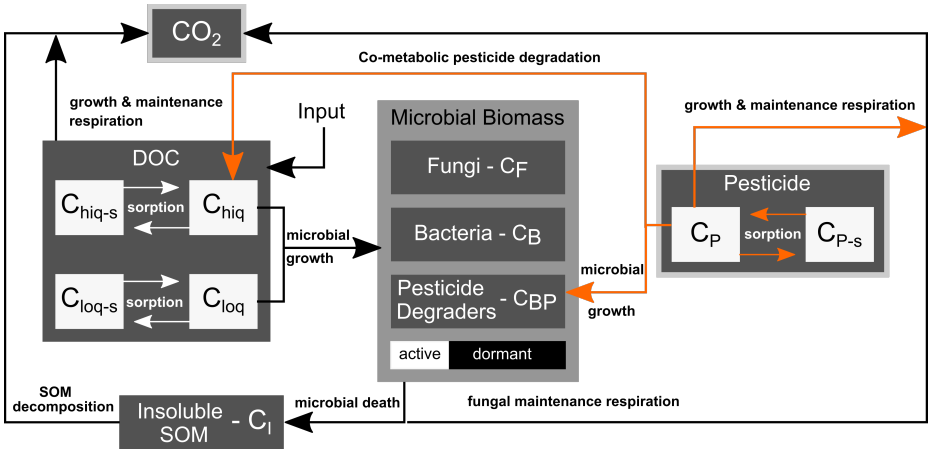


Figure 5.1 Diagram of the PECCAD ODE model (Pagel et al. 2014). Boxes symbolize C pools and arrows indicate C fluxes. The system can be observed at different resolutions: (i) using information on all available data including the dynamics of functional genes (dark gray), or (ii) bulk biomass (gray) along with measurements of dissolved organic C (DOC), insoluble soil organic matter (C_I), total pesticide (C_P + C_{P-s}) and heterotrophic respiration (CO_2); (iii) only with input-output information on total pesticide and CO_2 (light gray). Fluxes directly related to pesticide degradation are colored in orange. Individual C pools in white boxes correspond to unobserved system components.

tion 2.2 (relabeling $K(\vartheta) \equiv J(p)$ in this chapter) was minimized using the standard Levenberg-Marquardt algorithm (Press et al. 2007) with logarithmically transformed parameter values. Local sensitivity equations for calculating the Hessian (Eq. 2.9) and right hand side of the geodesic ODE (Eq. 2.40) were solved numerically by the direct differential method (Eq. 2.16) Ensembles of parameter sets were generated by Markov Chain Monte Carlo (MCMC) importance sampling (Gutenkunst 2007; Section 2.3.2) from the posterior distribution (Eq. 2.3) with log-normal priors that restricted parameters to lie with 95% confidence within two orders of magnitude of the locally-inferred best fit values. An available python script (Transtrum 2016) was used to implement the geodesic equation (Eq. 2.40) for the MBAM (Section 2.4). Algebraic details of selected MBAM model reduction steps are presented in the results section. For the Morris Method (Section 2.3.1) as implemented in the MATLAB SAFE toolbox for global sensitivity analysis (Pianosi et al. 2015), lower and upper bounds for uniform Latin Hypercube sampling of model parameters were set to 50% and 200% of the best fit parameter values. 25,000 trajectories were evaluated, corresponding to 2.64 million model evaluations. The robustness of Morris pairs was checked by bootstrapping and convergence analysis. Following the GSA approach by Link et al. 2018, Morris mean μ_i^* and standard deviation σ_i^*

were restricted to the unit square by normalizing with the largest value observed:

$$\hat{\mu}_i^* = \frac{\mu_i^*}{\max_i \mu_i^*}, \qquad \hat{\sigma}_i^* = \frac{\sigma_i^*}{\max_i \sigma_i^*}\,. \tag{5.1}$$

The ℓ_2-norm of normalized Morris pairs was subsequently used to screen for model parameters that have a negligible effect on the cost metric.

The original PECCAD model was validated with a series of microcosm experiments on the degradation of the herbicide MCPA in soil (see Pagel et al. 2016 for details). Briefly, homogenized soil was filled to a height of 30 mm into stainless steel cylinders (diameter 56 mm, height 40 mm) and compacted to a bulk density of $1.2\,g/cm^3$. In one experimental treatment (MCPA), an MCPA solution was homogeneously added to the soil to obtain an average concentration of $53\,\mu g/g$. In a second treatment (MCPA + Litter), the same MCPA amendment was used, but a layer consisting of 0.5 g maize litter was added on top of the microcosms. Four replicated microcosms of each treatment were then destructively sampled after 4.9, 7.8, 10.0, 13.9 and 22.8 days in 0-1,1-2, 2-3, 3-4, 4-6, 6-10 and 10-20 mm layers. To obtain sufficient material for analyses, soil from associated layers of two soil cores was combined resulting finally in two experimental replicates. In this study, data on MCPA, extractable DOC, total organic C (TOC), microbial biomass (C_B), genetic abundances of bacteria (*16S rRNA* genes), fungi (ITS fragments) and specific MCPA degraders (*tfdA* genes) was averaged over the first five soil layers (0-6 mm) of the microcosms. Thus, for each experimental treatment 67 data points were used to calibrate the PECCAD ODE model and estimate 59 parameters.

5.3 Results

5.3.1 Sloppiness and Systematic Reduction of PECCAD ODE

Figure 5.2 shows the Hessian eigenvalues of PECCAD ODE at each stage of the reduction by the MBAM. The initial 59 parameter model is *sloppy* when fit to the full data set of the MCPA + Litter experiment. The eigenvalues are uniformly spaced over 22 orders of magnitude. Thirty-two unidentifiable parameters correspond to eigenvalues smaller than one, i.e. manifold widths smaller than the scale set by the experimental measurement uncertainty. In each model reduction step, the smallest eigenvalue is removed from the spectrum by the MBAM without affecting the cost function value.

Tailoring model complexity to the full observational data set reduces the dimension of the PECCAD ODE system as well as the number of parameters (Tables 5.1 and 5.2, M=10, N=27). ODEs for physiological state indices of bacteria and specific pesticide de-

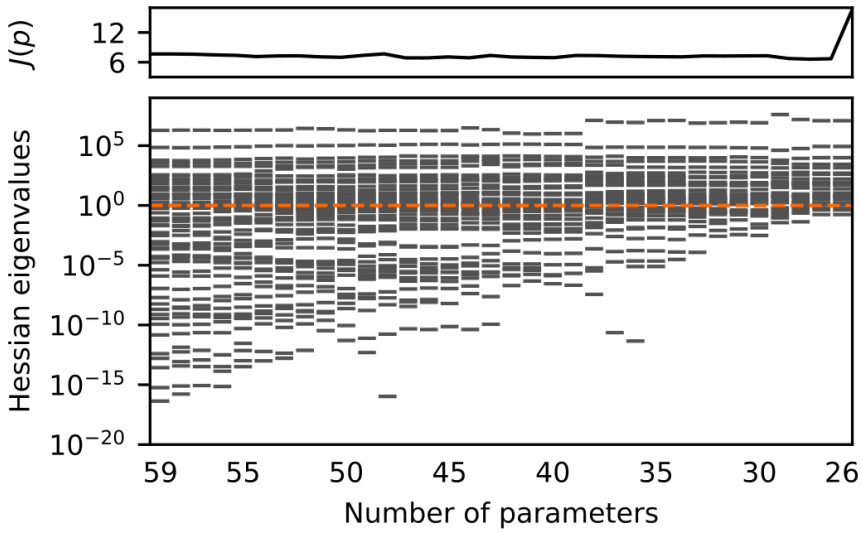


Figure 5.2 Tailoring model complexity of PECCAD ODE to functional gene data from the Litter + MCPA experiment. The MBAM removes one parameter at a time until the remaining parameters are identifiable to a given tolerance of 1/e (orange dashed line). Shown on top of the reduction spectra is the value of the cost function $J(p)$ during the iteration. The full observational data set identifies a 27 parameter model that fits the data equally well.

graders are transformed into algebraic equations that can be substituted into the original equation system. Nine effective parameters, fungal kinetic parameters, substrate uptake efficiencies and sorption coefficients govern the time evolution of the remaining C pools. The effective parameters are expressed in terms of nonlinear combinations of the original biokinetic parameters. Except for substrate uptake coefficients, only fungal parameters can be uniquely identified from the given data set. Fungal parameters related to the specific death rate are not constrained by data, but marginally important for model performance, i.e. they cannot be removed from the model without changing the value of the cost function (MBAM step 32 in Fig. 5.2). Except for the specific death rates of bacteria and fungi, all biokinetic functions originally formulated as multi-substrate Monod kinetics (Fig. 2.1) are sufficiently described by linear, rather than saturating functions of the substrate concentration.

Table 5.1 Governing differential equations and parameters of the reduced PECCAD ODE model (M=10, N=27) corresponding to the full MCPA + Litter data set.

C stock	Differential equation	Renormalized parameters [unit]	
Bacteria $[\frac{mg\,C}{g}]$	$\frac{dC_B}{dt} = \tilde{r}_B C_B \left(\vartheta_{22} C_{hiq} + \frac{\vartheta_{23} C_{loq}}{1+\vartheta_{24} C_{hiq}} - 1\right)$	$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}}$ $[\frac{g}{mg\,C}]$ $\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}}$ $[\frac{g}{mg\,C}]$ $\vartheta_{24} = \frac{k_{B,hiq}}{\mu_{max-B}}$ $[\frac{g}{mg\,C}]$	(1)
Fungi $[\frac{mg\,C}{g}]$	$\frac{dC_F}{dt} = r_F C_F \left(k_{F,loq} C_{loq} - (1 - Y_{r-F}) a_{max-F}\right) - r_F C_F \left(\frac{a_{max-F}}{1 + K_{a-F,hiq} C_{hiq} + K_{a-F,loq} C_{loq}}\right)$		(2)
Specific pesticide degraders $[\frac{mg\,C}{g}]$	$\frac{dC_{BP}}{dt} = \tilde{r}_{BP} C_{BP} \left(C_{loq}\right)$		(3)
Rescaled physiological state index of bacteria $[\frac{1}{d}]$	$\tilde{r}_B = \vartheta_{21} C_{loq}$	$\vartheta_{21} = \frac{a_{max-B}}{k_{r-B,loq}}$ $[\frac{g}{mg\,C}]$	(4)
Physiological state index of fungi $[-]$	$\frac{dr_F}{dt} = k_{F,loq} C_{loq} \left(\frac{C_{hiq}/k_{r-F,hiq}}{1+C_{hiq}/k_{r-F,hiq}} - r_F\right)$		(5)
Rescaled physiological state index of specific pesticide degraders $[\frac{1}{d}]$	$\tilde{r}_{BP} = \vartheta_{11} C_{hiq} + \vartheta_{12} C_{loq}$	$\vartheta_{11} = \frac{k_{B,loq}}{k_{r-B,hiq}}$ $[\frac{g^2}{(mg\,C)^2 d}]$ $\vartheta_{12} = \frac{k_{BP,loq}}{k_{r-BP,loq}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(6)
hiq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{hiq}}{dt} = I_{hiq}(t) + \tilde{r}_B C_B \left(-\frac{1}{Y_{s-B,hiq}} \vartheta_{22} C_{hiq}\right) + r_F C_F \left(Y_{R-F,P} \vartheta_{25} C_P C_{loq}\right)$	$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}}$ $[\frac{g}{mg\,C}]$ $\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(7)
Sorbed phase	$C_{hiq-s} = \frac{\rho_B}{\theta} K_{d-hiq} C_{hiq}$		(8)
loq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{loq}}{dt} = I_{loq}(t) + \tilde{r}_B C_B \left(-\frac{1}{Y_{s-B,loq}} \frac{\vartheta_{23} C_{loq}}{1+\vartheta_{24} C_{hiq}}\right) + r_F C_F \left(-\frac{1}{Y_{s-F,loq}} k_{F,loq} C_{loq}\right) + \tilde{r}_{BP} C_{BP} \left(-\frac{1}{Y_{s-BP,loq}} C_{loq}\right)$	$\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}}$ $[\frac{g}{mg\,C}]$ $\vartheta_{24} = \frac{k_{B,hiq}}{\mu_{max-B}}$ $[\frac{g}{mg\,C}]$	(9)

Sorbed phase

$$C_{loq-s} = \frac{\rho_B}{\theta} K_{d-loq} C_{loq} \tag{10}$$

Insoluble soil organic matter C $[\frac{mg\,C}{g}]$

$$\frac{dC_I}{dt} = \tilde{r}_B C_B + r_F C_F \left(\frac{a_{max-F}}{1 + K_{a-F,hiq} C_{hiq} + K_{a-F,loq} C_{loq}} \right) \tag{11}$$

Pesticide C $[\frac{mg\,C}{g}]$

$$\frac{dC_P}{dt} = -r_F C_F \vartheta_{25} C_{loq} C_P \qquad \vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}] \tag{12}$$

Sorbed phase

$$C_{P-s} = K_{F-P} \left(C_P \frac{1000 \rho_B M_P}{9 \theta M_c} \right)^{n_{F-P}} \tag{13}$$

CO_2 $[\frac{mg\,C}{g}]$

$$\begin{aligned} \frac{dCO_2}{dt} = & \frac{1 - Y_{s-B,hiq}}{Y_{s-B,hiq}} \tilde{r}_B C_B \vartheta_{22} C_{hiq} \\ & + \frac{1 - Y_{s-B,loq}}{Y_{s-B,loq}} \tilde{r}_B C_B \left(\frac{\vartheta_{23} C_{loq}}{1 + \vartheta_{24} C_{loq}} \right) \\ & + \frac{1 - Y_{s-BP,loq}}{Y_{s-BP,loq}} \tilde{r}_{BP} C_{BP} C_{loq} \\ & + \frac{1 - Y_{s-F,loq}}{Y_{s-F,loq}} r_F C_F k_{F,loq} C_{loq} \\ & + (1 - Y_{R-F,P}) r_F C_F \vartheta_{25} C_{loq} C_P \end{aligned} \tag{14}$$

$$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$$

$$\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$$

$$\vartheta_{24} = \frac{k_{B,hiq}}{\mu_{max-B}} \quad [\frac{g}{mg\,C}]$$

$$\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}]$$

Table 5.2 Model parameter symbols, descriptions, values of optimal parameters of the reduced PECCAD ODE model (M=10, N=27) calibrated on the data of two experimental treatments (MCPA + Litter, MCPA), 95% highest posterior density intervals (HDI) and units.

Symbol	Description	MCPA + Litter	HDI	MCPA	HDI	Unit
$\mathrm{a_{max-F}}$	Maximal specific death rate of fungi	43.03	$[6.80, 125.8]$	1.76	$[0.90, 5.06]$	$\left[\frac{1}{d}\right]$
$\mathrm{K_{a-F,hiq}}$	Inhibition coefficient of fungal death rate in response to hiq DOC	3772.04	$[382.79, 4.03e^3]$	78.39	$[17.82, 189.75]$	$\left[\frac{g}{mg\,C}\right]$
$\mathrm{K_{a-F,loq}}$	Inhibition coefficient of fungal death rate in response to loq DOC	133.80	$[21.55, 970.46]$	26.25	$[3.07, 132.17]$	$\left[\frac{g}{mg\,C}\right]$
$\mathrm{K_{d-hiq}}$	Linear sorption coefficient of hiq DOC	1.80	$[0.046, 3.63]$	$1e{-}3$	$[0.1, 3]e^{-3}$	$\left[\frac{mm^3}{mg}\right]$
$\mathrm{K_{d-loq}}$	Linear sorption coefficient of loq DOC	36.84	$[7.46, 70.29]$	280.18	$[86.94, 3.6e^5]$	$[-]$
$\mathrm{k_{F,loq}}$	loq DOC growth substrate affinity coefficient of fungi	32.74	$[11.09, 38.30]$	14.40	$[8.98, 19.61]$	$\left[\frac{g}{mg\,C\,d}\right]$
$\mathrm{k_{r-F,hiq}}$	Inhibition coefficient of fungal activity in response to hiq DOC	$3.5e^{-5}$	$[0.03, 1.4]e^{-4}$	$5.2e^{-3}$	$[3, 10]e^{-3}$	$\left[\frac{mg\,C}{g}\right]$
$\vartheta_{11} = \frac{k_{B,loq}}{k_{r-B,hiq}}$	Effective activity response of bacterial pesticide degraders in response to hiq DOC	156.97	$[0.012, 365.98]$	0.09	$[1e^{-4}, 0.20]$	$\left[\frac{g^2}{(mg\,C)^2 d}\right]$
$\vartheta_{12} = \frac{k_{BP,loq}}{k_{r-BP,loq}}$	Effective activity response of bacterial pesticide degraders to hiq DOC	6.01	$[5.045, 9.25]$	0.03	$[1e^{-4}, 0.12]$	$\left[\frac{g^2}{(mg\,C)^2 d}\right]$
$\vartheta_{21} = \frac{a_{max-B}}{k_{r-B,loq}}$	Effective activity response of bacteria to loq DOC	2.60	$[0.27, 4.91]$	1.83	$[1.50, 11.65]$	$\left[\frac{g}{mg\,C}\right]$
$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}}$	Effective hiq DOC uptake kinetic constant of bacteria	396.93	$[63.32, 2.11e^3]$	96.78	$[25.04, 80.83]$	$\left[\frac{g}{mg\,C}\right]$
$\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}}$	Renormalized loq DOC growth substrate affinity coefficient of bacteria	10.24	$[8.39, 17.36]$	0.06	$[1e^{-4}, 83.134]$	$\left[\frac{g}{mg\,C}\right]$
$\vartheta_{24} = \frac{k_{B,hiq}}{\mu_{max-B}}$	Effective inhibition coefficient of growth rate of bacteria in response to hiq DOC	1000.20	$[0.92, 2.08]e^3$	0.54	$[3e^{-3}, 2.4e^3]$	$\left[\frac{g}{mg\,C}\right]$

$\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}}$	Effective pesticide decomposition rate	$2.9e^6$	$[2.41, 6.93]e^6$	$4.8e^6$	$[3.7, 8.5]e^6$	$\left[\frac{g^2}{(mgC)^2 d}\right]$
$\mathrm{Y_{L,hiq}}$	Fraction of the decomposed hiq litter transferred to soil	0.36	$[0.10, 0.89]$	0.35	$[0.09, 0.49]$	$[-]$
$\mathrm{Y_{L,loq}}$	Fraction of the decomposed loq litter transferred to soil	0.89	$[0.80, 1.00]$	0.88	$[0.83, 0.98]$	$[-]$
$\mathrm{Y_{r-B}}$	Efficiency of insoluble organic matter decomposition by bacteria and bacterial pesticide degraders	0.75	$[0.56, 0.98]$	0.75	$[0.60, 0.86]$	$[-]$
$\mathrm{Y_{r-F}}$	Efficiency of insoluble organic matter decomposition by fungi 0.99	0.99	$[0.98, 1.0]$	0.99	$[0.993, 0.998]$	$[-]$
$\mathrm{Y_{R-F,P}}$	Efficiency of co-metabolic pesticide transformation by fungi	0.45	$[0.20, 0.86]$	0.83	$[0.81, 1.0]$	$[-]$
$\mathrm{Y_{s-B,hiq}}$	Substrate uptake efficiency of hiq DOC by bacteria	0.34	$[0.14, 0.44]$	0.28	$[0.35, 0.83]$	$[-]$
$\mathrm{Y_{s-B,loq}}$	Substrate uptake efficiency of loq DOC by bacteria	0.54	$[0.25, 0.62]$	0.63	$[0.62, 0.99]$	$[-]$
$\mathrm{Y_{s-BP,loq}}$	Substrate uptake efficiency of loq DOC by bacterial pesticide degraders	0.21	$[0.03, 0.75]$	0.21	$[0.05, 0.39]$	$[-]$
$\mathrm{Y_{s-F,loq}}$	Substrate uptake efficiency of loq DOC by fungi	0.92	$[0.85, 0.98]$	1.0	$[0.93, 1.0]$	$[-]$
$\mathrm{r_{F0}}$	Initial physiological state index of fungi	0.46	$[0.15, 0.81]$	$3.4e^{-3}$	$[0.1, 7]e^{-3}$	$[-]$

Model Simplification and Parameter Limits

In the following, parameter limits as defined by the geodesic equation ($\lim_{\tau \to \tau_b} p(\tau) = 0(\infty)$, where τ denotes the affine parameterization of the geodesic (Eq. 2.40) and τ_b denotes a manifold boundary) are simply referred to as $p \to 0(\infty)$.

Biokinetic functions of PECCAD ODE can be removed if the numerator of a rational rate expression in the original model (Table A12) approaches zero at a manifold boundary. As a result of *discarding limits* of this type, 12 out of 22 processes describing substrate dependent maintenance, growth, death and decomposition rates of specific functional microbial pools can be removed from the model without affecting the model performance.

Limits are less obvious when multiple parameters approach extreme values at the same rate. In these cases, *emergent* finite parameter combinations correspond to expressions such as ∞/∞, $0/0$, $0 \cdot \infty$ or $\infty - \infty$.

As an illustration of different types of limiting processes, consider, e.g., the following ODE of specific pesticide degrader C:

$$\frac{dC_{BP}}{dt} = r_{BP} C_{BP} \left(\mu_{BP,P} + \mu_{BP,hiq} + \mu_{BP,loq} - a_{BP} \right) . \tag{5.2}$$

The microbial pool changes through growth ($\mu_{BP,P}$, $\mu_{BP,hiq}$, $\mu_{BP,loq}$) and death (a_{BP}) and depends on the physiological state index of specific pesticide degraders which is a dynamic variable (r_{BP}, Blagodatsky and Richter 1998). Growth is possible on C_P, C_{hiq} and C_{loq}. Simultaneous utilization of growth substrates is accordingly modeled in terms of multi-substrate Monod kinetics (Fig. 2.1), where μ_{max-BP} is a maximum specific growth rate and $k_{BP,i}$, $i \in \{P, hiq, loq\}$ denote substrate specific affinity constants of bacterial pesticide degraders:

$$\mu_{BP,i} = \frac{\mu_{max-BP} k_{BP,i} C_i}{\mu_{max-BP} + k_{BP,loq} C_{loq} + k_{BP,hiq} C_{hiq} + k_{BP,P} C_P} . \tag{5.3}$$

The death rate (a_{BP}) is mediated by substrate availability in order to simulate increased microbial decay at low substrate concentrations, where a_{max-BP} likewise denotes a maximum specific death rate and $K_{a-BP,i}$, $i \in \{P, hiq, loq\}$ are substrate specific inhibition coefficients of microbial death:

$$a_{BP} = \frac{a_{max-BP}}{1 + K_{a-BP,loq} C_{loq} + K_{a-BP,hiq} C_{hiq} + K_{a-BP,P} C_P} . \tag{5.4}$$

Here, *discarding limits* correspond to $k_{BP,P} \to 0$, $k_{BP,hiq} \to 0$ and $K_{a-BP,P} \to 0$. That is, the time evolution of the specific degrader pool does not explicitly depend on the pesticide concentration C_P in the system:

$$\frac{dC_{BP}}{dt} = r_{BP} C_{BP} \left(\mu_{BP,loq} - \tilde{a}_{BP} \right) , \tag{5.5}$$

$$\tilde{a}_{BP} = \frac{a_{max-BP}}{1 + K_{a-BP,loq} C_{loq} + K_{a-BP,hiq} C_{hiq}} . \tag{5.6}$$

Additionally, *rescaling limits*, a_{max-BP}, $K_{a-BP,hiq}$, $K_{a-BP,loq} \to \infty$ allow to derive two effective finite parameter expressions
$\vartheta_1 = a_{max-BP}/K_{a-BP,loq}$, $\vartheta_2 = K_{a-BP,hiq}/K_{a-BP,loq}$, which control the effective specific

death rate $\tilde{a}_{BP}$. The rescaled expression becomes:

$$\tilde{a}_{BP} = \frac{a_{max-BP}}{K_{a-BP,loq}} \cdot \frac{1}{\frac{1}{K_{a-BP,loq}} + C_{loq} + \frac{K_{a-BP,hiq}}{K_{a-BP,loq}} C_{hiq}} \rightarrow \frac{\vartheta_1}{C_{loq} + \vartheta_2 C_{hiq}} \,. \tag{5.7}$$

Singular limits leading to steady-state approximations usually require evaluation of more than a single biokinetic term on the right hand side of the equation system. The following identified *singular limit* involves five ordinary differential equations of the system and six parameter limits, $\vartheta_1, \mu_{max-BP}, k_{BP,loq}, k_{r-BP,hiq}$, $k_{r-BP,loq}, m_{max-BP} \rightarrow \infty$:

$$\frac{dC_{BP}}{dt} = r_{BP} C_{BP} \left(\mu_{BP,loq} - \tilde{a}_{BP}\right) \tag{5.8}$$

$$\frac{dr_{BP}}{dt} = \mu_{BP,loq} \left(\Phi_{BP} - r_{BP}\right), \tag{5.9}$$

$$\Phi_{BP} = \frac{C_{hiq}/k_{r-BP,hiq} + C_{loq}/k_{r-BP,loq}}{1 + C_{hiq}/k_{r-BP,hiq} + C_{loq}/k_{r-BP,loq}} \tag{5.10}$$

$$\frac{dC_I}{dt} \propto r_{BP} C_{BP} \tilde{a}_{BP} \tag{5.11}$$

$$\frac{dC_{hiq}}{dt} \propto r_{BP} C_{BP} m_{max-BP} \tag{5.12}$$

$$\frac{dC_{loq}}{dt} \propto r_{BP} C_{BP} m_{max-BP} \tag{5.13}$$

Here, Φ_{BP} is a limiting factor of activity increase and m_{max-BP} is the maximum specific maintenance rate of bacterial pesticide degraders. First, if $\mu_{max-BP}, k_{BP,loq}, k_{r-BP,hiq}$, $k_{r-BP,loq} \rightarrow \infty$, then $r_{BP} \rightarrow 0$. Since $\vartheta_1, m_{max-BP} \rightarrow \infty$ at the same time, it follows that C_I, C_{hiq} and C_{loq} become infinitely sensitive to changes in r_{BP} and the combination $r_{BP}\vartheta_1$ or $r_{BP} m_{max-BP}$ remains finite. The equations then read

$$\frac{dC_{BP}}{dt} = \tilde{r}_{BP} C_{BP} \left(\tilde{\mu}_{BP,loq} - \frac{1}{C_{loq} + \vartheta_2 C_{hiq}}\right), \tag{5.14}$$

$$\tilde{\mu}_{BP,loq} = \frac{\vartheta_4 C_{loq}}{\vartheta_5 + C_{loq}} \tag{5.15}$$

$$\frac{1}{\mu_{max-BP}} \frac{d\tilde{r}_{BP}}{dt} = \frac{\tilde{\mu}_{BP,loq}}{\vartheta_4} \left(\Phi_{BP}\vartheta_1 - \tilde{r}_{BP}\right) \tag{5.16}$$

$$\frac{dC_I}{dt} \propto \tilde{r}_{BP} C_{BP} \frac{1}{C_{loq} + \vartheta_2 C_{hiq}} \tag{5.17}$$

$$\frac{dC_{hiq}}{dt} \propto \tilde{r}_{BP} C_{BP} \vartheta_8 \tag{5.18}$$

$$\frac{dC_{loq}}{dt} \propto \tilde{r}_{BP} C_{BP} \vartheta_8 \,, \tag{5.19}$$

where $\vartheta_4 = \frac{\mu_{max-BP}}{\vartheta_1}$, $\vartheta_5 = \frac{\mu_{max-BP}}{k_{BP,loq}}$, $\vartheta_8 = \frac{m_{max-BP}}{\vartheta_1}$. I chose to define a renormalized variable $\tilde{r}_{BP} = r_{BP}\vartheta_1$, thereby removing information about the absolute scale of the activity level of bacterial pesticide degraders. From Eq. 5.16 it is obvious that $\varepsilon = 1/\mu_{max-BP}$ is a small parameter that separates the timescale of the renormalized variable $\tilde{r}_{BP}$. Evaluating this limit has the biological interpretation as a natural steady-state limit in which the physiological state of bacterial pesticide degraders is determined by the scaled substrate response function Φ_{BP} (Eq. 5.10; Blagodatsky and Richter 1998). The ODE for the physiological state index of specific pesticide degraders (Eq. 5.16) is transformed into an algebraic equation that can be substituted into the original ODE system:

$$\tilde{r}_{BP} = \Phi_{BP}\vartheta_1 \,. \tag{5.20}$$

As a result, *singular limits* identified via manifold boundaries decrease the dimension of the ODE system.

Finally, *interpolating limits* dictate the order of a reaction rate. The limit in which both Monod constants become infinite together ($\vartheta_4, \vartheta_5 \to \infty$ in Eq. 5.15) identifies a linear rate with *emergent* rate constant $\vartheta_9 = \vartheta_4/\vartheta_5$:

$$\tilde{\mu}_{BP,loq} = \frac{\vartheta_4 C_{loq}}{\vartheta_5 + C_{loq}} = \frac{\vartheta_4}{\vartheta_5}\frac{C_{loq}}{1 + \frac{C_{loq}}{\vartheta_5}} \to \vartheta_9 C_{loq} \,. \tag{5.21}$$

The alternative limit $\vartheta_5 \to 0$ would have corresponded to a saturating approximation of Monod kinetics (cf. Fig. 2.1).

Model Performance

By design, the full and reduced models give a reasonable fit within the expected variance of experimental uncertainties to data from the MCPA + Litter treatment (Fig. 5.3, black dots) with cost function value $J_{full} = 7.6$ for the full model and $J_{reduced} = 6.7$ for the reduced model. The first MBAM iteration resulted in a better (local) optimum. Time series generated from the reduced model (Fig. 5.3, red dashed lines) for fungal C, specific degrader C, DOC and $CO_2 - C$ are almost an exact match with the corresponding time series of the full model (Fig. 5.3, dark gray solid lines). For MCPA, the lag phase of MCPA degradation is reflected slightly better than by the reduced model. In contrast to the full model, steady-state conditions for the TOC pool are not yet reached after 25 days in the reduced model. Furthermore, bacterial C dynamics notably differ between the full and the reduced models.

The fit of the reduced model to MCPA treatment data (Fig. 5.3, green dotted line and

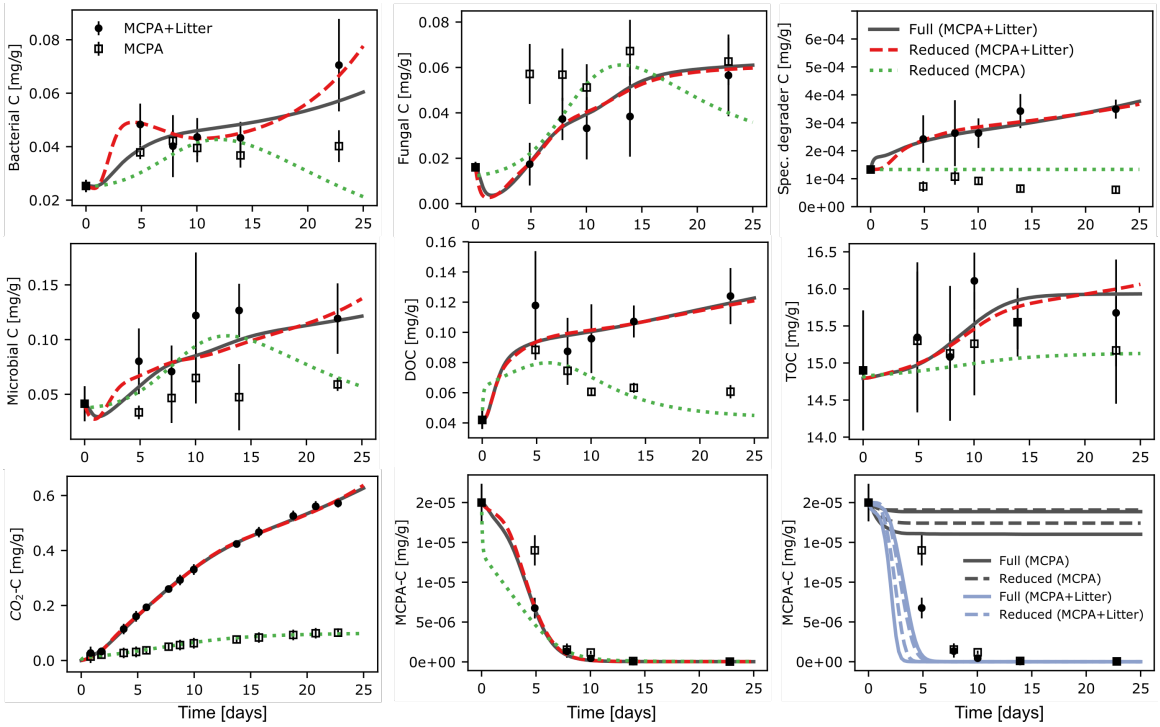


Figure 5.3 Model calibration and prediction. Full (dark gray solid lines) and reduced (red dashed lines) models give a reasonable fit to data from the MCPA + Litter treatment (circles, $J_{full} = 7.6$, $J_{reduced} = 6.7$). The fit to MCPA treatment data (squares) of the reduced model (green dotted lines) is worse ($J_{reduced} = 83.5$), because microbial dynamics are not fully captured. 95% confidence intervals for MCPA predictions between experimental treatments are shown in the bottom right panel. Predictions between experimental treatments of the full and reduced models (predicted data set in round brackets) derived from an MCMC parameter ensemble are well-constrained given the observed MCPA-C range, but do not match experimental observations.

black squares) is worse ($J_{reduced,MCPA} = 83.5$). The reduced model neither captures the dynamics of specific MCPA degrading bacteria nor the decelerated degradation of MCPA in the initial phase of the experiment without litter addition. When fit to MCPA treatment data, the eigenvalue spectrum of the reduced 27 parameter model broadens again and information on seven model parameters is lost (Fig. 5.4).

Model predictions of MCPA dynamics for shifted boundary conditions according to different experimental treatments are shown in the bottom right panel of Fig. 5.3. Full and reduced models were both calibrated on MCPA + Litter data and used for prediction of the observed MCPA dynamics in the experiment without litter addition (MCPA) and the other way around. The 95% confidence intervals for model predictions derived

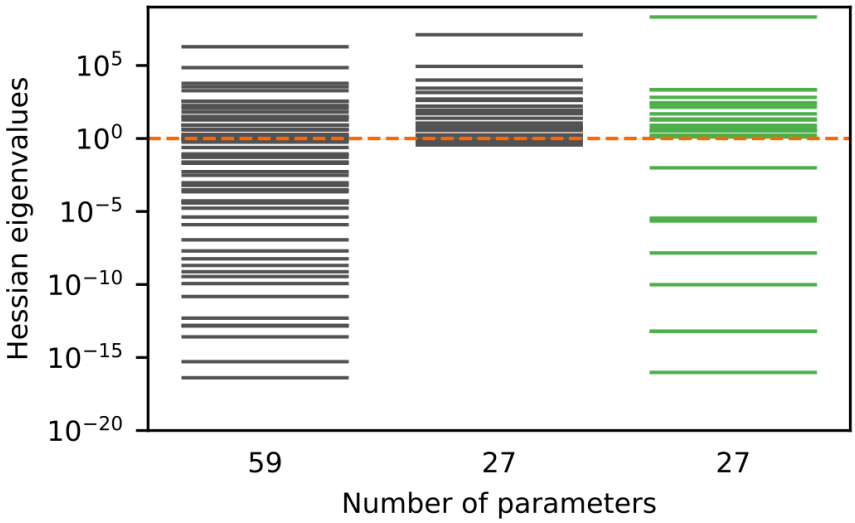


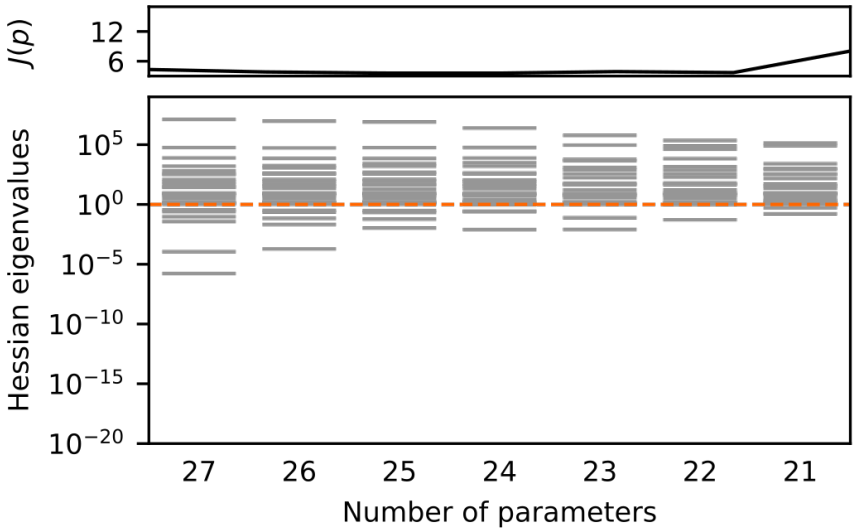
Figure 5.4 Quantifying information in MCPA data. Upon fitting the reduced 27 parameter model to the MCPA experimental treatment the Hessian eigenvalue spectrum broadens again and information on 7 model parameters is lost.

from a Bayesian ensemble of the full and reduced models for both data sets are informative (the limits span less than 15% of the total MCPA-C concentration range), but do not match experimental observations. When calibrated on MCPA + Litter data, reduced and full models predict MCPA persistence in soil after four days when no litter is added to the system in contrast to the observed complete dissipation in the experiment. Conversely, when calibrated on MCPA data, both models over-predict the acceleration of MCPA degradation in the presence of additional litter C input into the system.

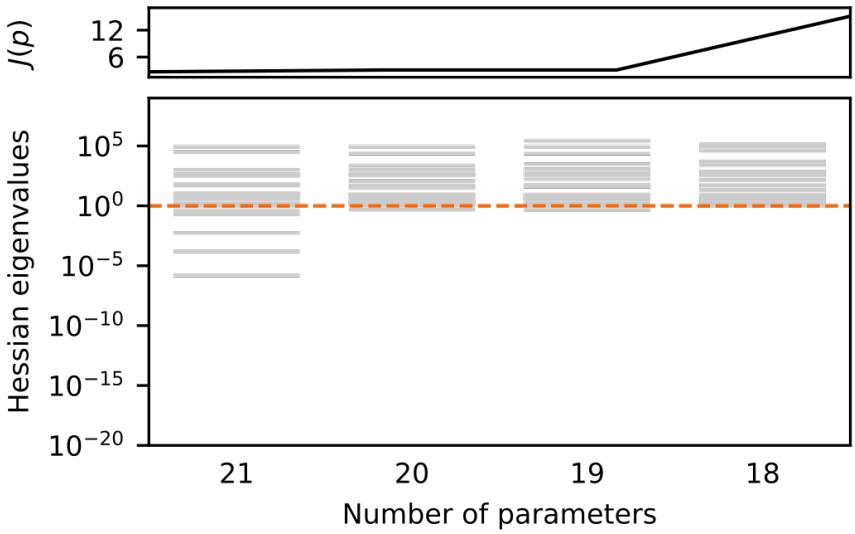
Impact of Data Availability on Model Reduction

Using the reduced 27 parameter model (Table 5.1 and Fig. 5.2) as a starting point, the effect of coarsening the observations from functional gene measurements to bulk microbial biomass and further to MCPA concentration and heterotrophic respiration is depicted in Fig. 5.5ab. After coarsening to bulk biomass (Fig. 5.5a), six eigenvalues become significantly smaller than unity. The parameter limits correspond to $\vartheta_{24}, \vartheta_{11}, \vartheta_{12}$, $a_{max-F}, K_{a-F,hiq}, K_{a-F,loq} \to 0$. The resulting *discarding* limits render the microbial death rate linear and remove the fungal death rate as well as the dependence of C cycling on the dynamics of specific pesticide degraders (Table 5.3, M=7, N=21).

Coarsening the observations further to system input-output relations (only MCPA and $CO_2 - C$) identifies an 18 parameter model that describes the dynamics of MCPA degra-



(a) Bulk data



(b) Input-Output data

Figure 5.5 Tailoring model complexity of PECCAD ODE to coarse-grained data sets from the Litter + MCPA experiment. The MBAM removes one parameter at a time until the remaining parameters are identifiable to a given tolerance of 1/e (orange dashed line). Shown on top of the reduction spectra is the value of the cost function $J(p)$ during the iteration. (a) Coarsening observations from functional gene measurements (Fig. 5.2) to bulk microbial biomass identifies a 21 parameter model governed by 7 ODEs. (b) Observing only MCPA and heterotrophic soil respiration identifies an 18 parameter ODE of dimension 6.

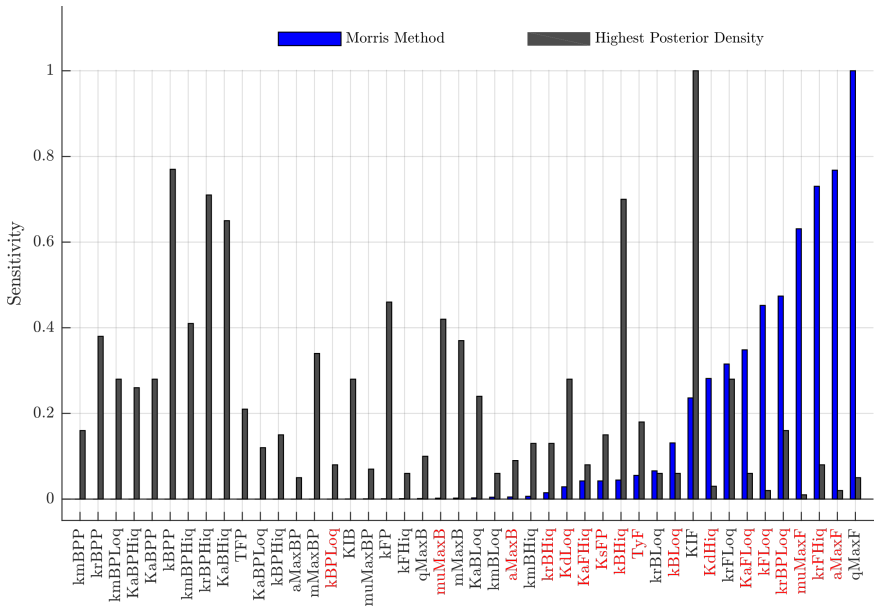


Figure 5.6 Comparison of MBAM results to sampling-based sensitivity metrics. Axis labels highlighted in red are PECCAD ODE kinetic model parameters that were identified as relevant by the MBAM. The normalized ℓ_2-norm of 25,000 Morris pairs (blue, sorted in ascending order) identifies a 20 parameter subset that influences the model performance metric. The 95% highest posterior density of 20 parameters spans more than 20% of their prior range after Bayesian model calibration. Overall, the screening results agree with the MBAM. Surprisingly, the most influential parameter q_{max-F} identified by the Morris method and with significant reduction in highest posterior density is not part of the reduced model.

dation and heterotrophic respiration (Fig. 5.5b). Here, another *discarding* limit ($\vartheta_{21} \to 0$) corresponds to a steady-state limit that fixes the insoluble organic matter pool (C_I) to its initial value (Table 5.4, M=6, N=18).

5.3.2 Global Sensitivity Analysis

The Morris procedure was applied to the parameters of the full PECCAD ODE model in order to compare sampling-based criteria for factor fixing and screening in global sensitivity applications to the results of the MBAM. Fig. 5.6 shows the ℓ_2-norm of the normalized Morris mean $\hat{\mu}^*$ and standard deviation $\hat{\sigma}^*$ ($\ell_2 = \sqrt{\hat{\mu}^{*2} + \hat{\sigma}^{*2}}$) and the relative reduction in highest posterior density of parameter values derived from the Bayesian model calibration (Table A13). Out of 20 model parameters with non-trivial ℓ_2-norm > 0.01 , 15 parameters agree with the MBAM results. In contrast to MBAM, $q_{max-F}, k_{r-F,loq}, K_{I_F}, k_{r-B,loq}$ and $k_{m-B,hiq}$ have significant effects on the goodness-of-fit metric. Two parameters that are essential to the reduced model (μ_{max-B} and $k_{BP,loq}$) were not identified by the Morris

Method. The results of the Bayesian model analysis shows that the 95% highest posterior density of 20 parameters still span more than 20% of their prior range after optimization. A clear cutoff that defines identifiability does not exist.

Table 5.3 Governing differential equations and parameters of the reduced PECCAD ODE model (M=7, N=21) corresponding to bulk measurements in the MCPA + Litter data set.

C stock	Differential equation	Renormalized parameters [unit]	
Bacteria $[\frac{mg\,C}{g}]$	$\frac{dC_B}{dt} = \tilde{r}_B C_B \left(\vartheta_{22} C_{hiq} + \vartheta_{23} C_{loq} - 1\right)$	$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$ $\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$	(1)
Fungi $[\frac{mg\,C}{g}]$	$\frac{dC_F}{dt} = r_F C_F \left(k_{F,loq} C_{loq}\right)$		(2)
Specific pesticide degraders $[\frac{mg\,C}{g}]$	$\frac{dC_{BP}}{dt} = 0$		(3)
Rescaled physiological state index of bacteria $[\frac{1}{d}]$	$\tilde{r}_B = \vartheta_{21} C_{loq}$	$\vartheta_{21} = \frac{a_{max-B}}{k_{r-B,loq}} \quad [\frac{g}{mg\,C}]$	(4)
Physiological state index of fungi $[-]$	$\frac{dr_F}{dt} = k_{F,loq} C_{loq} \left(\frac{C_{hiq}/k_{r-F,hiq}}{1+C_{hiq}/k_{r-F,hiq}} - r_F\right)$		(5)
Rescaled physiological state index of specific pesticide degraders $[\frac{1}{d}]$	$\tilde{r}_{BP} = 0$		(6)
hiq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{hiq}}{dt} = I_{hiq}(t) + \tilde{r}_B C_B \left(-\frac{1}{Y_{s-B,hiq}} \vartheta_{22} C_{hiq}\right) + r_F C_F \left(Y_{R-F,P} \vartheta_{25} C_P C_{loq}\right)$	$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$ $\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}]$	(7)
Sorbed phase	$C_{hiq-s} = \frac{\rho_B}{\theta} K_{d-hiq} C_{hiq}$		(8)
loq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{loq}}{dt} = I_{loq}(t) + \tilde{r}_B C_B \left(-\frac{1}{Y_{s-B,loq}} \vartheta_{23} C_{loq}\right) + r_F C_F \left(-\frac{1}{Y_{s-F,loq}} k_{F,loq} C_{loq}\right)$	$\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$	(9)
Sorbed phase	$C_{loq-s} = \frac{\rho_B}{\theta} K_{d-loq} C_{loq}$		(10)
Insoluble soil organic matter C $[\frac{mg\,C}{g}]$	$\frac{dC_I}{dt} = \tilde{r}_B C_B$		(11)
Pesticide C $[\frac{mg\,C}{g}]$	$\frac{dC_P}{dt} = -r_F C_F \vartheta_{25} C_{loq} C_P$	$\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}]$	(12)
Sorbed phase	$C_{P-s} = K_{F-P} \left(C_P \frac{1000 \rho_B M_P}{9\theta M_c}\right)^{n_{F-P}}$		(13)

CO_2 $[\frac{mg\,C}{g}]$	$\frac{dCO_2}{dt} = \frac{1-Y_{s-B,hiq}}{Y_{s-B,hiq}}\tilde{r}_B C_B \vartheta_{22} C_{hiq} + \frac{1-Y_{s-B,loq}}{Y_{s-B,loq}}\tilde{r}_B C_B \left(\vartheta_{23} C_{loq}\right) + \frac{1-Y_{s-F,loq}}{Y_{s-F,loq}} r_F C_F k_{F,loq} C_{loq} + (1-Y_{R-F,P}) r_F C_F \vartheta_{25} C_{loq} C_P$	$\vartheta_{22} = \frac{k_{B,hiq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$ $\vartheta_{23} = \frac{k_{B,loq}}{a_{max-B}} \quad [\frac{g}{mg\,C}]$ $\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}]$

(14)

Table 5.4 Carbon stocks, governing differential equations and renormalized parameters of the reduced PECCAD ODE model (M=6, N=18) corresponding to input-output observations of the MCPA + Litter experiment.

C stock	Differential equation	Renormalized parameters	
Bacteria $[\frac{mg\,C}{g}]$	$\frac{dC_B}{dt} = \tilde{r}_B C_B \left(\vartheta_{22} C_{hiq} - 1\right)$	$\vartheta_{22} = \frac{k_{B,loq}}{k_{r-B,loq}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(1)
Fungi $[\frac{mg\,C}{g}]$	$\frac{dC_F}{dt} = r_F C_F \left(k_{F,loq} C_{loq}\right)$		(2)
Specific pesticide degraders $[\frac{mg\,C}{g}]$	$\frac{dC_{BP}}{dt} = 0$		(3)
Rescaled physiological state index of bacteria $[\frac{mg\,C}{g}]$	$\tilde{r}_B = C_{loq}$		(4)
Physiological state index of fungi $[-]$	$\frac{dr_F}{dt} = k_{F,loq} C_{loq} \left(\frac{\tilde{C}_{hiq}/\vartheta_{29}}{1+\tilde{C}_{hiq}/\vartheta_{29}} - r_F\right)$	$\vartheta_{29} = \frac{k_{B,hiq} k_{r-F,loq}}{k_{r-B,loq}}$ $[\frac{g}{(mg\,C) d}]$	(5)
Rescaled physiological state index of specific pesticide degraders $[\frac{1}{d}]$	$\tilde{r}_{BP} = 0$		(6)
Rescaled *hiq* dissolved organic C $[\frac{g}{mg\,C d}]$	$\tilde{C}_{hiq} = \frac{Y_{s-B,hiq}}{\tilde{r}_B C_B}\left(I_{hiq}(t)\right) + \frac{Y_{s-B,hiq}}{\tilde{r}_B C_B}\left(r_F C_F \left(Y_{R-F,P} \vartheta_{25} C_P C_{loq}\right)\right)$	$\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(7)
Sorbed phase	$C_{hiq-s} = 0$		(8)
loq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{loq}}{dt} = I_{loq}(t) + \tilde{r}_B C_B \left(-\frac{1}{Y_{s-B,loq}} \vartheta_{28} C_{loq}\right) + r_F C_F \left(-\frac{1}{Y_{s-F,loq}} k_{F,loq} C_{loq}\right)$	$\vartheta_{28} = \frac{k_{B,loq}}{k_{r-B,loq}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(9)
Sorbed phase	$C_{loq-s} = \frac{\rho_B}{\theta} K_{d-loq} C_{loq}$		(10)
Insoluble soil organic matter C $[\frac{mg\,C}{g}]$	$\frac{dC_I}{dt} = 0$		(11)
Pesticide C $[\frac{mg\,C}{g}]$	$\frac{dC_P}{dt} = -r_F C_F \vartheta_{25} C_{loq} C_P$	$\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}}$ $[\frac{g^2}{(mg\,C)^2 d}]$	(12)
Sorbed phase	$C_{P-s} = K_{F-P}\left(C_P \frac{1000 \rho_B M_P}{9 \theta M_c}\right)^{n_{F-P}}$		(13)

CO_2 $[\frac{mg\,C}{g}]$	$\frac{dCO_2}{dt} = \frac{1-Y_{s-B,hiq}}{Y_{s-B,hiq}}\tilde{r}_B C_B \tilde{C}_{hiq} + \frac{1-Y_{s-B,loq}}{Y_{s-B,loq}}\tilde{r}_B C_B \left(\vartheta_{28} C_{loq}\right) + \frac{1-Y_{s-F,loq}}{Y_{s-F,loq}} r_F C_F k_{F,loq} C_{loq} + (1-Y_{R-F,P}) r_F C_F \vartheta_{25} C_{loq} C_P$	$\vartheta_{28} = \frac{k_{B,loq}}{k_{r-B,loq}} \quad [\frac{g^2}{(mg\,C)^2 d}]$ $\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}} \quad [\frac{g^2}{(mg\,C)^2 d}]$	(14)

5.4 Discussion

Possible regulation mechanisms of MCPA degradation have been extensively discussed in Poll et al. 2010; Pagel et al. 2016. Based on inverse modeling with PECCAD, Pagel et al. 2016 concluded that fungal dynamics probably play a crucial role for matter cycling in the detritusphere (i.e. the soil influenced by litter). They found that MCPA degradation in soil was likely predominantly regulated by co-metabolic degradation via litter-stimulated fungal growth. Uncertainty in this statement stems from the fact that their results were based on the interpretation of single parameter values with high uncertainty (Pareto ranges for 26 out of 59 biokinetic parameters were equal to their respective prior range after optimization). Systematic model reduction of PECCAD ODE by the MBAM (Table 5.1) reflects the reported dominance of co-metabolic over direct MCPA degradation in the original study by Pagel et al. 2016.

In the reduced PECCAD ODE model, it is obvious that MCPA degradation is controlled by litter C input. The fraction of C_{loq} transported into the system stimulates fungal growth. The fungal specific growth rate is simply a linear function of substrate concentration. The emergent microbial "control knob" of MCPA degradation is an effective "renormalized" rate ($\vartheta_{25} = \frac{k_{F,loq}}{T_{y-F} K_{s-F,P}}$) that depends on the substrate affinity of fungi to low quality C ($k_{F,loq}$) and co-metabolic pesticide transformation kinetics ($K_{s-F,P}$), as well as the capacity of fungi to transform MCPA into high quality C substrates for growth (T_{y-F}). Growth of bacteria is the only microbial process that contributes to degradation of high quality C, while specific pesticide degraders exclusively grow on low quality C. Their dynamics are otherwise detached from C cycling and MCPA turnover. On the contrary, it has repeatedly been empirically observed that activity and abundance of the population of specific MCPA degraders increase in the presence of their preferred growth substrate (Poll et al. 2010; Saleh et al. 2016). However, the experimental data of the Litter + MCPA treatment (total abundance of functional genes of specific MCPA degraders) probably does not contain enough information on the dynamics of specific pesticide degraders to accordingly constrain the model structure of the reduced PECCAD ODE model. Most biokinetic parameters related to direct MCPA degradation are irrelevant for model behavior. As a result, the simulated specific degrader dynamics by the reduced model do not match experimental observations across different treatments and the model fails in forecasting MCPA dynamics. Similar conclusions on the representation of microbial dynamics based on computationally expensive multiobjective calibration of multimodal data (Wöhling et al. 2013) were drawn in the original study (Pagel et al. 2016). Data-driven model reduction highlights the need to increase the structural complexity of the PECCAD ODE model or to refine the process description of direct pesticide degradation.

Optimized parameter values of the reduced model (Table 5.2) categorize fungi as copiotrophic organisms (Fierer et al. 2007) in both experimental treatments. The value of the activity inhibition coefficient $k_{r-F,hiq}$ at the lower bound of its physiological range shows that fungi respond extremely fast to supply of high quality C. This finding is in line with Ingwersen et al. 2008; Pagel et al. 2016 who also reported a high sensitivity of fungal activity to low concentrations of easily degradable substrates. Early onset of activity was interpreted as being stimulated by intermediate degradation products of high quality C-derived extracellular enzymes that induce enzyme production (Allison et al. 2010*b*). Concurrently, optimized values for substrate efficiencies (Y_s) of fungi are much higher than for bacteria and specific pesticide degraders. Mortality rates could only be estimated from MCPA treatment data (Table 5.2). The value of the maximum specific death rate of fungi ($a_{max-F} = 1.76\,1/d$) is close to the values reported for copiotrophic organisms in other studies (Zelenev et al. 2005; Monga et al. 2014).

Overall, the results of the PECCAD ODE reduction show that the mathematical specification of biokinetic functions in terms of multi-substrate Monod kinetics is an unnecessarily strong assumption. It is mostly sufficient to model C cycling rates as linear functions of substrate concentration. Steady-state modeling of bacterial and specific pesticide degrader activity is a nontrivial result of the reduction process. The feedback structure between microbial populations and C sources of different quality is preserved under the full observational data set. As model reduction is capable of directly altering model structure, this emphasizes the notion of soil as a complex dynamical system.

Model structure is deemphasized later in the reduction process, upon coarse-graining of observations. The assumption of consistency for the SOM pool prevalent in the early soil C modeling literature (Jenkinson 1990; Parton et al. 1987; Carvalhais et al. 2008) is only supported for input-output observations (Table 5.4).

The results of the MBAM are in good agreement with global sensitivity measures derived from the Morris method and Bayesian model calibration (Fig. 5.6). In contrast to global sensitivity methods, the MBAM is deterministic and does not rely on sampling of the parameter space. The sensitivity results are easier to translate into model simplifications, because parameters that do not affect model output are identified iteratively. In contrast, factor fixing will reduce the number of model parameters that have to be considered, e.g., in refined calibrations (e.g., Van Werkhoven et al. 2009), but it only makes the model conceptually simpler if the zero value is in the range of parameter variations that do not affect model output. Furthermore, Latin-Hypercube sampling of the parameter space

of the PECCAD ODE model proved difficult, because the rate of failure of numerical integration was 76% on average. Three times as many function evaluations as expected from the theoretical limit (Section 2.3.1) were hence required to achieve the stated number of sensitivity evaluations. The MBAM on the other hand is not fully automatic and still requires manual intervention in its current implementation.

Chapter 6

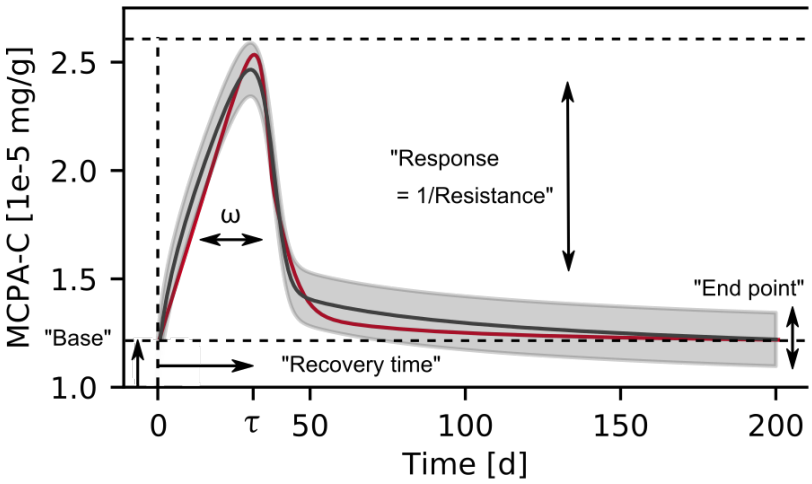
Emergent Minimal Models of Biogeochemical System Behavior

6.1 Introduction

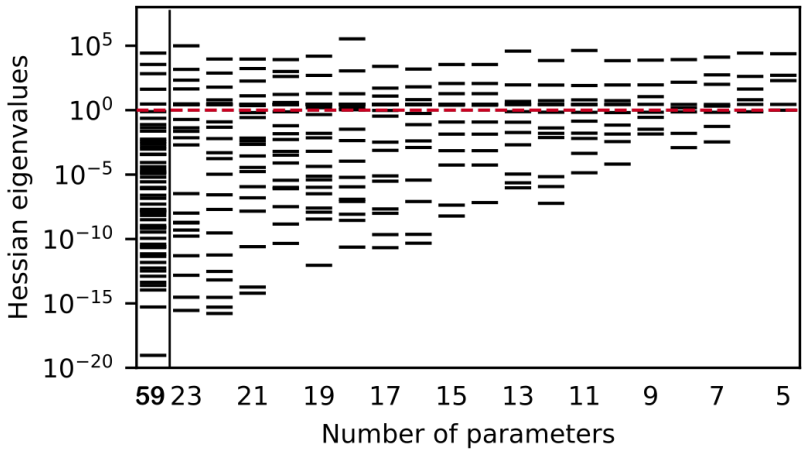
The fate of pesticides in soil is regulated by the complex interplay between microbial dynamics and physicochemical processes (Pagel et al. 2016). The complexity of individual soil processes stands, however, in stark contrast to the surprisingly simple *emergent* behavior exhibited by the associated systems. In contrast to the many parameters that define the underlying biogeochemical models (Section 2.1), the system response can often be described by a few phenomenological parameters. In order to illustrate this discrepancy, the *dynamic stability* of a soil system after a chemical press disturbance (Schaeffer et al. 2016) is analyzed in this chapter[1].

Fig. 6.1a shows the soil system response to continuous pesticide input (a step-function input) at time zero. The system is *resilient*, i.e. it shows the ability to return to its initial state after a disturbance. The system response function is characterized by five degrees of freedom that can be linked more generally to ecological notions of ecosystem resilience and stability (Hastings et al. 2018). According to the terminology in the conceptual framework for soils (Schaeffer et al. 2016, Table 1), the dynamic stability is defined as the time that a polluted soil needs to respond and in this case recover to a stable state after stress impact. The faster the *recovery time*, the higher the dynamic stability. Phenomenologically, the recovery time is determined by the time to achieve maximal *response* (τ) and the width of the response peak (ω). The *resistance* of soils in turn is defined as the reciprocal of the response to a given stress intensity. It is given by the height of the response peak at time τ. Finally, the system behavior is characterized by the difference between final (*end point*) and initial (*base*) states and the vertical position of the base line value. A minimal phenomenological model of pesticide stabilization would have to contain parameters that span these degrees of freedom.

[1] This is not to be confused with a stability analysis from a dynamical systems point of view in which stability properties of dynamical systems to small perturbations of initial conditions are analyzed (Kuznetsov 2013; Strogatz 2018). For applied stability analysis in soil systems, see, e.g., Manzoni and Porporato 2007; Wang et al. 2014*b*.



(a) Phenomenology



(b) Model reduction

Figure 6.1 Degrees of freedom that characterize pesticide stabilization. (a) One realization of full PECCAD ODE (M=12, N=59; black line) was used as a basis for model reduction. The gray uncertainty band sets the tolerance for allowed deviations from the simulated MCPA-C time course. The system response can be described by five phenomenological parameters: (1) the time to achieve maximal response (τ), (2) the width of the response peak (ω), (3) the height of the peak, (4) the difference between the base and end points, and (5) the vertical position of the base line value. The reduced PECCAD ODE model (M=5, N=5; red line) achieves stabilization within the given tolerance. (b) PECCAD ODE eigenvalues during MBAM iterations (truncated). The five parameters of reduced PECCAD ODE that correspond to the five degrees of freedom of the pesticide stabilization curve are identifiable at each stage of the reduction (to a given tolerance, the five eigenvalues above the red dashed line). The MBAM removes the parameter corresponding to the smallest eigenvalue in each iteration.

6.2 Implementation

One realization of PECCAD ODE (Table A14) that exhibits pesticide stabilization was used as a starting point for model reduction by the MBAM. The initial conditions for all carbon pools except for soluble MCPA were set to zero. The model output curve for pesticide was sampled at 100 time points in order to simulate almost continuous data. Normally-distributed noise with standard deviation equal to 5% of the maximum pesticide concentration was added to the model output in order to set the allowed tolerance for deviations from the desired stabilizing behavior.

The technical implementation of PECCAD ODE and the MBAM was identical to Chapter 5.

6.3 Results

Minimal Mechanisms of Pesticide Stabilization

Figure 6.1b shows the Hessian eigenvalues before the first and during the last 18 MBAM iterations. The full PECCAD ODE model with 59 parameters is *sloppy* even for densely-sampled data with minimal measurement error. At each stage of the reduction the MBAM would identify a five parameter model with parameters that correspond to eigenvalues larger than unity. The number of parameters in the reduced model (Table 6.1) is in agreement with the number of degrees of freedom that phenomenologically describe the pesticide stabilization curve (Fig. 6.1a). The dimension of the ODE system is reduced to five. The system response still depends on the dynamics of specific pesticide degraders C_{BP}, their physiological state index r_{BP}, total pesticide $C_P + C_{P-s}$ and low quality dissolved organic carbon C_{loq} (the insoluble soil organic matter pool C_I is only needed in order to close the mass balance). The five parameters that govern model behavior are given by the substrate uptake efficiency of bacterial pesticide degraders for pesticide, $Y_{s-BP,P}$, and low quality carbon, $Y_{s-BP,loq}$, the ratio of the maximum specific growth rate to the maximum specific death rate of bacterial pesticide degraders, $\vartheta_{28} = \mu_{max-BP}/a_{max-BP}$, the ratio of their respective growth substrate affinities to pesticide and low quality carbon, $\vartheta_6 = k_{BP,P}/k_{BP,loq}$ and the scaled Freundlich coefficient of the pesticide sorption isotherm, $\tilde{K}_{F-P} = K_{F-P}C_P(t=0)$.

The five parameters of the reduced model can be connected to the degrees of freedom of the stabilization curve through a local sensitivity analysis (Fig. 6.2). To this end, parameter values were increased by 150% of their respective baseline values. Perturbing ϑ_6 or $Y_{s-BP,loq}$ (Fig. 6.2a and b) increases or decreases the final equilibrium value, whilst keep-

Table 6.1 Carbon stocks and governing differential equations of the minimal PECCAD ODE model (M=5, N=5) for pesticide stabilization.

C stock	Differential equation
Specific pesticide degraders $[\frac{mg\,C}{g}]$	$\frac{dC_{BP}}{dt} = r_{BP}C_{BP}\left(\frac{\vartheta_{28}C_P}{\vartheta_6 C_{loq}+C_P} + \frac{\vartheta_{28}C_{loq}}{C_{loq}+C_P/\vartheta_6} - 1\right)$
Physiological state index of specific pesticide degraders $[-]$	$\frac{r_{BP}}{dt} = \left(\frac{\vartheta_{28}C_P}{\vartheta_6 C_{loq}+C_P} + \frac{\vartheta_{28}C_{loq}}{C_{loq}+C_P/\vartheta_6}\right)\cdot(1-r_{BP})$
loq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{loq}}{dt} = I_{loq} - r_{BP}C_{BP}\left(\frac{1}{Y_{s-BP,loq}}\frac{\vartheta_{28}C_{loq}}{C_{loq}+C_P/\vartheta_6}\right)$
Insoluble soil organic matter C $[\frac{mg\,C}{g}]$	$\frac{dC_I}{dt} = r_{BP}C_{BP}$
Pesticide C $[\frac{mg\,C}{g}]$	$\frac{d(C_P+C_{P-s})}{dt} = I_P - r_{BP}C_{BP}\left(\frac{1}{Y_{s-BP,P}}\frac{\vartheta_{28}C_P}{\vartheta_6 C_{loq}+C_P}\right)$
Sorbed phase	$C_{P-s} = K_{F-P}C_P(0)\cdot\frac{1000\rho_B M_P}{9\theta M_c}$

ing the response constant. Varying ϑ_{28} (Fig. 6.2c) has the most dramatic effect on the degrees of freedom. It changes the response and dynamic stability of the system through τ and ω, whilst lowering the final steady state value. Changing $Y_{s-BP,P}$ affects the width of the response peak and raises the final steady state value. Varying a combination of ϑ_6 and $Y_{s-BP,loq}$ leads to full system recovery. Increasing ϑ_{28} and $Y_{s-BP,P}$ simultaneously is the "microbial control knob" for the resistance of the soil system to a chemical press disturbance. Increasing the renormalized Freundlich coefficient moves the system into a state with historically higher pollution (Fig. 6.2e).

6.4 Discussion

PECCAD ODE captures key small-scale biogeochemical mechanisms of pesticide degradation. Through model reduction with the MBAM, the mechanisms can be explicitly linked to high-level governing principles of pesticide stabilization.

The toy example demonstrates that it is possible to construct minimal effective process descriptions that remain expressed in terms of mechanistic parameters from complicated models. By design, the effective five parameter model can fit artificial data generated from the full model. Structural errors introduced by the simplification are indistinguish-

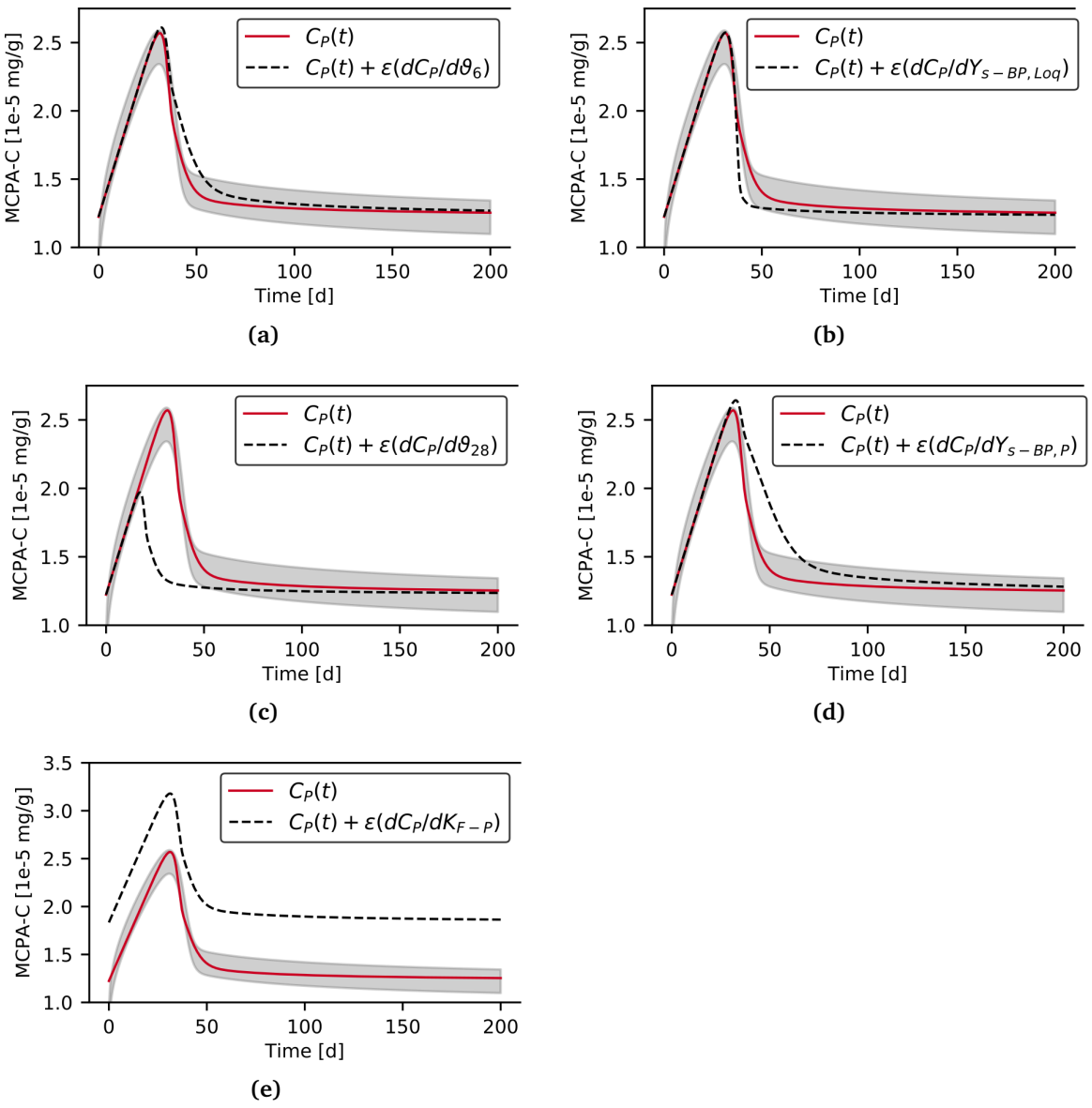


Figure 6.2 Parameter sensitivities in the minimal model for pesticide stabilization (Table 6.1). An effective five parameter model can capture the response of the full model to a step input of MCPA-C. Increasing the five parameters by 150% from their base values ($\vartheta_6 = 3.81$, $\vartheta_{28} = 1.53$, $Y_{s-BP,loq} = 0.08$, $Y_{s-BP,P} = 0.10$, $K_{F-P} = 0.72\,mm^3/mg$) moves the model response outside of the experimental uncertainty band (gray). Combinations of parameters span the five phenomenological degrees of freedom of the system response curve (Fig. 6.1a).

able from noise. If the remaining few microbial parameters (maximum specific growth and death rates, substrate affinities and uptake efficiencies) as well as the Freundlich sorption coefficient of pesticide could be soil-specifically measured, it would be possible to test the predictive power of the minimal model without the need for a model calibration step.

Since the effective process description *emerges* from the underlying biogeochemical mechanisms, the resulting minimal model could be used in larger-scale models with greater confidence than a priori imposed top-down expressions for pesticide stabilization. However, for the approach to be valid at all for up-scaling the *universality* of the *emergent* minimal model structure has to be assessed[2]. Theoretically, universality for pesticide stabilization could be assessed by characterizing the viable space for which the PECCAD ODE model exhibits experimentally observed behavior. This is best done by a global, followed by a local exploration of parameter space as described in Zamora-Sillero et al. 2011. Statistical models in the viable parameter space volume then correspond to models that have "stabilization behavior" as a point somewhere on their model manifold. The question would be whether MBAM applications for the different models converge to the same *emergent* minimal model structures.
Alternatively, the uncertainty band around the full model output could be increased. The resulting minimal model will likely have fewer effective parameters and still exhibit stabilization. However, the reduced model will be unable to exactly mirror the curvature of the full model output. Applicability would depend on the actual measurement uncertainty and the variance in observed pesticide concentrations.

Microbial response functions that capture the macro-scale sensitivity of heterotrophic respiration or enzymatic decomposition to changes in soil moisture or temperature are bound to replace empirical response functions (Sierra et al. 2015*b*) in biogeochemical models. However, mechanistic descriptions that exist in the literature are still too complex to be embedded into ecosystem models (Manzoni et al. 2016). Renormalizing complex biogeochemical models into simpler models with only a few effective parameters is a promising way to construct microbial response functions in a mechanistically informed manner. The approach could supersede ad-hoc methods for constructing minimal process models, e.g., in the context of modeling microbal responses to the rewetting of dry soils (Manzoni et al. 2014; Hawkes et al. 2017; Brangari et al. 2018).

[2]For example for chemical diffusion, the effective diffusion constant depends on the underlying stochastic particle model and the observed data (Machta et al. 2013; Transtrum et al. 2015), but the same functional form for the diffusion equation is always recovered in the continuum limit.

Chapter 7

Synthesis

As outlined in the introduction of this thesis, a tension between small-scale complexity and *emergent* simplicity of observables in biogeochemical systems exists that needs to be resolved in order for the associated models to be predictive. I used a novel information theoretic framework developed in the Sethna lab at Cornell University to address obstacles to an efficient and meaningful use of mechanistic biogeochemical modeling. In this final chapter, the insights of four different case studies (Chapters 3 to 6) are synthesized.

The Origin of Equifinality in Biogeochemical Modeling

For five biogeochemical models of different complexity from the biogeochemical literature, the ubiquity of *equifinality* has been traced down to global *sloppiness* and the associated *hyper-ribbon* structure of the model prediction manifolds (Chapter 3).

The problems with parameter inference from experimental data are not merely a consequence of how the biogeochemical models have been parameterized, suggesting that modelers are simply consistently unfortunate in choosing, for example, the correct units or rates versus time constants to describe system behavior. Instead the *equifinality* problem originates in the *conceptual uncertainty* in the model itself, given the information content of the data. Combining many empirical laws, e.g., variants of Monod-type equations (Section 3.3.3), into complex biogeochemical models leads to *sloppiness*. *Sloppiness* can be removed, because the boundary complex of the model manifold corresponds to natural, mechanistically-meaningful limits of biogeochemical models: *interpolating* limits dictate the order of a biogeochemical reaction rate, *singular limits* lead to a separation of timescales in the system and *discarding* limits remove irrelevant pathways in the model (Sections 4.3.3 and 5.3.1). Moreover, *sloppiness* appears to be independent of the *measurement uncertainty* in the data. The degree of *sloppiness* of the PECCAD ODE model was the same for synthetic continuous versus real data (Figs. 3.1, 5.5 and 6.1).

Sloppiness might be avoided by imposing tighter prior constraints on the parameter values. However, in situ estimates of microbial parameters are still highly uncertain and their values can potentially span many orders of magnitude (Pagel et al. 2014, Online Resource 1, Allison 2017, Table 1). For the analyzed biogeochemical models it would seem artificial to make strong assumptions about parameter values and model structures

in order to avoid *sloppiness*. Furthermore, any experimental protocol designed to remove *sloppiness* would be intimately tied to the uncertain model structure that is to be falsified. As a result, there will likely always be potential for *equifinality* in biogeochemical modeling.

The results of this thesis show that the appropriateness of biogeochemical model assumptions can be assessed despite *equifinality*, given any information content of the data. *Sloppiness* as the origin of *equifinality* can be systematically eliminated by a geodesic-following algorithm that exploits the *sloppy* structure of parameter space (Transtrum and Qiu 2014).

Appropriate Biogeochemical Model Complexity

One important criterion to improve mechanistic modeling frameworks for complex systems is the ability to adequately encode model complexity (e.g., Schöniger et al. 2014; Getz et al. 2018; Höge et al. 2018). For *sloppy* biogeochemical models the complexity is simply given by the effective dimensionality of the model manifold. Information criteria that score against metric properties of the full *sloppy* model would overestimate model complexity in biogeochemical model selection scenarios (LaMont and Wiggins 2015; Mattingly et al. 2018).

Model complexity is an integral part of the principle of parsimony (Occam's razor) popularized in the statement that "the simplest solution is usually the correct one". For predictive modeling it is routinely stated that the best model must reflect "both criteria that make up the Bayesian tradeoff, goodness-of-fit and parsimony (i.e., relatively small variance in its predictions)" (Schöniger et al. 2015). It is believed that "non-identifiable parameters often lead to imprecise model predictions." (Maiwald et al. 2016). For *sloppy* biogeochemical models the latter interpretation of Occam's law as a bias-variance tradeoff is false. The conceptual uncertainty in biogeochemical models leads to large parameter uncertainties, thereby hampering system understanding, but it does not affect model prediction uncertainty. For example, the full and reduced PECCAD ODE models fit the data equally well and make statistically almost indistinguishable model predictions with low variance (Fig. 5.3). Full and reduced PECCAD ODE are effectively the "same" model, because the model prediction manifold of full PECCAD ODE has a hyper-ribbon structure, i.e. a low effective dimensionality much less than the number of model parameters. Predictions from ensemble fits can be precise despite large uncertainties in the *sloppy* parameter subspace (Fig. 3.6c). Conversely, knowing all but one parameter a priori might give little information about the model behavior, if the precision of this parameter measurement were to exceed the extent of the *stiffest* direction in parameter

space (Fig. 3.6a). This justifies the use of purely data-constrained parameter estimates to improve predictive biogeochemical model performance (e.g., Hararuk et al. 2015).

The principle of parsimony for *sloppy* biogeochemical models is solely reflected in the emergent minimal model structures. The sole important implication for model performance is that parameters of the effective model are identifiable (Fig. 4.6 and Table 5.2). Fundamental mechanistic understanding from experimental studies that is otherwise obfuscated by model complexity can be established by model reduction. By systematically coarsening observations (Fig. 5.5) it is possible to assess which type of mechanistic information can be transferred across focal levels (Getz et al. 2018) at which the relevant research questions are being asked. For example, the fungal substrate affinity coefficient for growth on low quality carbon in PECCAD ODE is a parameter that remains relevant as the scale of observation grows (Tables 5.1, 5.3 and 5.4). Although I am aware of the fact that "there are no general laws for complexity"(Goldenfeld and Kadanoff 1999), empirical evidence points to the fact that it is mostly sufficient to describe complex soil systems with intricate feedback structures by linear effective degradation rates and equilibrium assumptions for the activity of a subset of microbial functional types and carbon pools (Table 5.1).

However, finding a process representation that is informative with regard to the predictions of interest is a nontrivial task (Fig. 5.3). Among many possible refinements for the representation of MCPA degradation in PECCAD ODE (WP3a, Collaborative Research Center 1253 CAMPOS funded by Deutsche Forschungsgemeinschaft (Grant Agreement SFB 1253/1 2017)) one solution might, e.g., consider the incorporation of energy-limited growth on carbon substrates. Again, different conceptualizations exist for energy-limited growth (Desmond-Le Quéméner and Bouchez 2014; LaRowe and Amend 2015). While this thesis shows that a procedure is in place that allows to construct effective parsimonious biogeochemical models, the problem of including the correct mechanistic processes in the first place, i.e. the "Landau step" of renormalization in the physics analogy (Nemenman 2017) is much more difficult. In mechanistic modeling frameworks that require inversion to generate the posterior distribution of model predictions, a sensitivity analysis will usually be needed to determine whether calibration and forecast data are driven by the same parameters. Any refinement that adds complexity to the original model would likely entail the necessity of a subsequent simplification step to recover parsimony. Model reduction by the MBAM (Transtrum and Qiu 2014) is especially suited to the task, because it directly reveals conceptual uncertainty in the original model formulation (Chapter 5).

Conclusion

The presented results suggest that information geometry provides a powerful approach to connect the inherent small-scale complexity of microbiological processes to the *emergent* behavior of biogeochemical systems. My thesis contributes to solutions for two of the biggest challenges in biogeochemical modeling: to assess the validity of model assumptions in light of *equifinality* and to construct mechanistic models that solve the inference problem from data. The Manifold Boundary Approximation Method (Transtrum and Qiu 2014) is a valuable addition to existing model-data integration frameworks in biogeochemical modeling that seek to maximize mechanistic system understanding. It remedies drawbacks of global sensitivity analysis methods which depend on the sampling of high-dimensional parameter spaces and the results of which are impossible to translate into the model simplifications presented in this thesis. Its potential to routinely assist commonly used global sensitivity schemes will hinge on the technical implementation. The computational overhead requires the integration of local sensitivity information with respect to all model parameters. This task is usually feasible for biogeochemical models and can be parallelized. The step that sets the barrier of entry quite high for general usage is the manual translation of parameter limits into analytical model simplifications. As a first step, my current research activities are intensely concentrating on finding an autonomous solution that complements the analysis of short-term substrate-induced heterotrophic respiration data. Through an appropriate redefinition of the statistical model, the method could be easily extended to include prior information on biogeochemical parameters in the future. Above all, my thesis highlights that potential applications of the applied approach throughout mechanistic biogeochemical modeling are truly manifold.

References

Abbena, E., S. Salamon, and A. Gray. 2017. Modern differential geometry of curves and surfaces with Mathematica. Chapman and Hall/CRC.

Adamson, M. and A. Y. Morozov. 2012. When can we trust our model predictions? Unearthing structural sensitivity in biological systems. Proc. R. Soc. A, page rspa20120500.

Addiscott, T. 2011. Emergence or self-organization? Look to the soil population. Communicative & Integrative Biology, **4**:469–470.

Allison, S. D. 2017. Building Predictive Models for Diverse Microbial Communities in Soil. In Microbial Biomass: A Paradigm Shift in Terrestrial Biogeochemistry, pages 141–166. World Scientific.

Allison, S. D., M. D. Wallenstein, and M. A. Bradford. 2010*a*. Soil-carbon response to warming dependent on microbial physiology. Nature Geoscience, **3**:336.

Allison, S. D., M. N. Weintraub, T. B. Gartner, and M. P. Waldrop. 2010*b*. Evolutionary-economic principles as regulators of soil enzyme production and ecosystem function. In Soil Enzymology, pages 229–243. Springer.

Amari, S.-i. and H. Nagaoka. 2007. Methods of Information Geometry, volume 191. American Mathematical Soc.

Anderson, J. and K. Domsch. 1978. A physiological method for the quantitative measurement of microbial biomass in soils. Soil Biology and Biochemistry, **10**:215–221.

Apgar, J. F., D. K. Witmer, F. M. White, and B. Tidor. 2010. Sloppy models, parameter uncertainty, and the role of experimental design. Molecular BioSystems, **6**:1890–1900.

Aslam, S., P. Benoit, F. Chabauty, V. Bergheaud, C. Geng, L. Vieublé-Gonod, and P. Garnier. 2014. Modelling the impacts of maize decomposition on glyphosate dynamics in mulch. European Journal of Soil Science, **65**:231–247.

Ay, N., J. Jost, H. Vân Lê, and L. Schwachhöfer. 2017. Information Geometry, volume 8. Springer.

Babey, T., L. Vieublé-Gonod, A. Rapaport, M. Pinheiro, P. Garnier, and J.-R. De Dreuzy. 2017. Spatiotemporal simulations of 2,4-D pesticide degradation by microorganisms in 3D soil-core experiments. Ecological Modelling, **344**:48–61.

Bælum, J., C. S. Jacobsen, and W. E. Holben. 2010. Comparison of 16S rRNA gene phylogeny and functional tfdA gene distribution in thirty-one different 2,4-dichlorophenoxyacetic acid and 4-chloro-2-methylphenoxyacetic acid degraders. Systematic and Applied Microbiology, **33**:67–70.

Bælum, J., M. H. Nicolaisen, W. E. Holben, B. W. Strobel, J. Sørensen, and C. S. Jacobsen. 2008. Direct analysis of tfdA gene expression by indigenous bacteria in phenoxy acid amended agricultural soil. The ISME Journal, **2**:677.

Ballantyne IV, F. and S. A. Billings. 2018. Model formulation of microbial CO2 production and efficiency can significantly influence short and long term soil C projections. The ISME journal, page 1.

Banitz, T., K. Johst, L. Y. Wick, S. Schamfuß, H. Harms, and K. Frank. 2013. Highways versus pipelines: contributions of two fungal transport mechanisms to efficient bioremediation. Environmental Microbiology Reports, **5**:211–218.

Bates, D. and D. Watts. 1988. Nonlinear regression analysis and its applications. Wiley series in probability and mathematical statistics, Wiley, New York [u.a.].

Baveye, P. C., W. Otten, A. Kravchenko, M. Balseiro-Romero, E. Beckers, M. Chalhoub, C. Darnault, T. Eickhorst, P. Garnier, S. Hapca, et al. 2018. Emergent properties of microbial activity in heterogeneous soil microenvironments: different research approaches are slowly converging, yet major challenges remain. Frontiers in Microbiology, **9**.

Beven, K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology, **320**:18–36.

Bezanson, J., A. Edelman, S. Karpinski, and V. B. Shah. 2017. Julia: A fresh approach to numerical computing. SIAM review, **59**:65–98.

Blagodatskaya, E. and Y. Kuzyakov. 2013. Active microorganisms in soil: critical review of estimation criteria and approaches. Soil Biology and Biochemistry, **67**:192–211.

Blagodatsky, S. and O. Richter. 1998. Microbial growth in soil and nitrogen turnover: a theoretical model considering the activity state of microorganisms. Soil Biology and Biochemistry, **30**:1743–1755.

Blagodatsky, S. and P. Smith. 2012. Soil physics meets soil biology: towards better mechanistic prediction of greenhouse gas emissions from soil. Soil Biology and Biochemistry, **47**:78–92.

Blagodatsky, S. A., O. Heinemeyer, and J. Richter. 2000. Estimating the active and total soil microbial biomass by kinetic respiration analysis. Biology and Fertility of Soils, **32**:73–81.

Bohner, G. and G. Venkataraman. 2017. Identifiability, reducibility, and adaptability in allosteric macromolecules. The Journal of General Physiology, **149**:547–560.

Brangari, A. C., S. Manzoni, and J. Rousk. 2018. Mechanistically modeling the microbial growth and respiration responses to rewetting dry soils. In EGU General Assembly Conference Abstracts, volume 20, page 9434.

Brown, K. S. and J. P. Sethna. 2003. Statistical mechanical approaches to models with many poorly known parameters. Physical Review E, **68**:021904.

Brown, P. N., A. C. Hindmarsh, and L. R. Petzold. 1994. Using Krylov methods in the solution of large-scale differential-algebraic systems. SIAM Journal on Scientific Computing, **15**:1467–1488.

Buchanan, M. 2015. Wheat from the chaff. Nature Physics, **11**:296.

Burbea, J. and C. R. Rao. 1982. Entropy differential metric, distance and divergence measures in probability spaces: A unified approach. Journal of Multivariate Analysis, **12**:575–596.

Calderhead, B. and M. Girolami. 2011. Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods. Interface Focus, page rsfs20110051.

Campolongo, F., J. Cariboni, and A. Saltelli. 2007. An effective screening design for sensitivity analysis of large models. Environmental Modelling & Software, **22**:1509–1518.

Campolongo, F., A. Saltelli, and J. Cariboni. 2011. From screening to quantitative sensitivity analysis. A unified approach. Computer Physics Communications, **182**:978–988.

Carvalhais, N., M. Reichstein, J. Seixas, G. J. Collatz, J. S. Pereira, P. Berbigier, A. Carrara, A. Granier, L. Montagnani, D. Papale, et al. 2008. Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval. Global Biogeochemical Cycles, **22**.

Casey, F. P., D. Baird, Q. Feng, R. N. Gutenkunst, J. J. Waterfall, C. R. Myers, K. S. Brown, R. A. Cerione, and J. P. Sethna. 2007. Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. IET Systems Biology, **1**:190–202.

Caticha, A. 2015. The basics of information geometry. In AIP Conference Proceedings, volume 1641, pages 15–26. AIP.

Cencov, N. N. 2000. Statistical decision rules and optimal inference. Number 53, American Mathematical Soc.

Chib, S. and E. Greenberg. 1995. Understanding the metropolis-hastings algorithm. The American Statistician, **49**:327–335.

Choe, Y. 2017. Information criterion for minimum cross-entropy model selection. arXiv preprint arXiv:1704.04315.

Colores, G. M., S. K. Schmidt, and M. C. Fisk. 1996. Estimating the biomass of microbial functional groups using rates of growth-related soil respiration. Soil Biology and Biochemistry, **28**:1569–1577.

Desmond-Le Quéméner, E. and T. Bouchez. 2014. A thermodynamic theory of microbial growth. The ISME Journal, **8**:1747.

Ditterich, F., C. Poll, H. Pagel, D. Babin, K. Smalla, M. A. Horn, T. Streck, and E. Kandeler. 2013. Succession of bacterial and fungal 4-chloro-2-methylphenoxyacetic acid degraders at the soil–litter interface. FEMS Microbiology Ecology, **86**:85–100.

Do Carmo, M. P. 2016. Differential Geometry of Curves and Surfaces: Revised and Updated Second Edition. Courier Dover Publications.

Dötsch, A., J. Severin, W. Alt, E. A. Galinski, and J.-U. Kreft. 2008. A mathematical model for growth and osmoregulation in halophilic bacteria. Microbiology, **154**:2956–2969.

Duan, Q., S. Sorooshian, and V. Gupta. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research, **28**:1015–1031.

Dungait, J. A., D. W. Hopkins, A. S. Gregory, and A. P. Whitmore. 2012. Soil organic matter turnover is governed by accessibility not recalcitrance. Global Change Biology, **18**:1781–1796.

Engl, H. W., C. Flamm, P. Kügler, J. Lu, S. Müller, and P. Schuster. 2009. Inverse problems in systems biology. Inverse Problems, **25**:123014.

Engl, H. W., M. Hanke, and A. Neubauer. 1996. Regularization of inverse problems, volume 375. Springer Science & Business Media.

Fick, A. 1855. Ueber Diffusion. Annalen der Physik, **170**:59–86.

Fierer, N., A. Barberán, and D. C. Laughlin. 2014. Seeing the forest for the genes: using metagenomics to infer the aggregated traits of microbial communities. Frontiers in Microbiology, **5**:614.

Fierer, N., M. A. Bradford, and R. B. Jackson. 2007. Toward an ecological classification of soil bacteria. Ecology, **88**:1354–1364.

Gábor, A. and J. R. Banga. 2015. Robust and efficient parameter estimation in dynamic models of biological systems. BMC Systems Biology, **9**:74.

Gauss, C. F. 1877. Allgemeine Lehrsätze in Beziehung auf die im verkehrten Verhältnisse des Quadrats der Entfernung wirkenden Anziehungs-und Abstossungs-Kräfte. In Werke, pages 195–242. Springer.

Georgiou, K., R. Z. Abramoff, J. Harte, W. J. Riley, and M. S. Torn. 2017. Microbial community-level regulation explains soil carbon responses to long-term litter manipulations. Nature Communications, **8**:1223.

Geris, L., D. Gomez-Cabrero, et al. 2016. Uncertainty in biology. Springer.

German, D. P., K. R. Marcelo, M. M. Stone, and S. D. Allison. 2012. The Michaelis–Menten kinetics of soil extracellular enzymes in response to temperature: a cross-latitudinal study. Global Change Biology, **18**:1468–1479.

Getz, W. M., C. R. Marshall, C. J. Carlson, L. Giuggioli, S. J. Ryan, S. S. Romañach, C. Boettiger, S. D. Chamberlain, L. Larsen, P. DOdorico, et al. 2018. Making ecological models adequate. Ecology Letters, **21**:153–166.

Gharasoo, M., F. Centler, P. Regnier, H. Harms, and M. Thullner. 2012. A reactive transport modeling approach to simulate biogeochemical processes in pore structures with pore-scale heterogeneities. Environmental Modelling & Software, **30**:102–114.

Goldenfeld, N. and L. P. Kadanoff. 1999. Simple lessons from complexity. Science, **284**:87–89.

Gómez, H. F., M. Hucka, S. M. Keating, G. Nudelman, D. Iber, and S. C. Sealfon. 2016. Moccasin: converting Matlab Ode models to Sbml. Bioinformatics, **32**:1905–1906.

Grimm, V. and S. F. Railsback. 2012. Pattern-oriented modelling: a multi-scopefor predictive systems ecology. Phil. Trans. R. Soc. B, **367**:298–310.

Gutenkunst, R. 2007. Sloppiness, modeling, and evolution in biochemical networks. Dissertation. Cornell University.

Gutenkunst, R. N., J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. 2007. Universally sloppy parameter sensitivities in systems biology models. PLoS Computational Biology, **3**:e189.

Hamamoto, S., P. Moldrup, K. Kawamoto, and T. Komatsu. 2010. Excluded-volume expansion of Archie's law for gas and solute diffusivities and electrical and thermal conductivities in variably saturated porous media. Water Resources Research, **46**.

Hararuk, O., M. J. Smith, and Y. Luo. 2015. Microbial models with data-driven parameters predict stronger soil carbon responses to climate change. Global change biology, **21**:2439–2453.

Hassani, S. 2013. Mathematical physics: a modern introduction to its foundations. Springer Science & Business Media.

Hastings, A., K. C. Abbott, K. Cuddington, T. Francis, G. Gellner, Y.-C. Lai, A. Morozov, S. Petrovskii, K. Scranton, and M. L. Zeeman. 2018. Transient phenomena in ecology. Science, **361**. URL `http://science.sciencemag.org/content/361/6406/eaat6412`.

Hawkes, C. V. and T. H. Keitt. 2015. Resilience vs. historical contingency in microbial responses to environmental change. Ecology Letters, **18**:612–625.

Hawkes, C. V., B. G. Waring, J. D. Rocca, and S. N. Kivlin. 2017. Historical climate controls soil respiration responses to current soil moisture. Proceedings of the National Academy of Sciences, **114**:6322–6327.

Hermans, T. 2017. Prediction-focused approaches: An opportunity for hydrology. Groundwater, **55**:683–687.

Höge, M., T. Wöhling, and W. Nowak. 2018. A Primer for Model Selection: The Decisive Role of Model Complexity. Water Resources Research.

Hug, L. A., B. J. Baker, K. Anantharaman, C. T. Brown, A. J. Probst, C. J. Castelle, C. N. Butterfield, A. W. Hernsdorf, Y. Amano, K. Ise, et al. 2016. A new view of the tree of life. Nature Microbiology, **1**:16048.

Hunt, C. A., A. Erdemir, W. W. Lytton, F. Mac Gabhann, E. A. Sander, M. K. Transtrum, and L. Mulugeta. 2018. The Spectrum of Mechanism-Oriented Models and Methods for Explanations of Biological Phenomena. Processes, **6**:56.

Ingwersen, J., C. Poll, T. Streck, and E. Kandeler. 2008. Micro-scale modelling of carbon turnover driven by microbial succession at a biogeochemical interface. Soil Biology and Biochemistry, **40**:864–878.

Iooss, B. and P. Lemaître. 2015. A review on global sensitivity analysis methods. In Uncertainty management in simulation-optimization of complex systems, pages 101–122. Springer.

Jenkinson, D. S. 1990. The turnover of organic carbon and nitrogen in soil. Phil. Trans. R. Soc. Lond. B, **329**:361–368.

Keenan, T. F., E. A. Davidson, J. W. Munger, and A. D. Richardson. 2013. Rate my data: quantifying the value of ecological data for the development of models of the terrestrial carbon cycle. Ecological Applications, **23**:273–286.

Kovárová-Kovar, K. and T. Egli. 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiology and Molecular Biology Reviews, **62**:646–666.

Kuehn, C. 2016. Multiple time scale dynamics, volume 1. Springer.

Kügler, P. 2016. Early afterdepolarizations with growing amplitudes via delayed subcritical Hopf bifurcations and unstable manifolds of saddle foci in cardiac action potential dynamics. PloS ONE, **11**:e0151178.

Kügler, P., M. Bulelzai, and A. H. Erhardt. 2017. Period doubling cascades of limit cycles in cardiac action potential models as precursors to chaotic early Afterdepolarizations. BMC Systems Biology, **11**:42.

Kuznetsov, Y. A. 2013. Elements of applied bifurcation theory, volume 112. Springer Science & Business Media.

Kuzyakov, Y. and E. Blagodatskaya. 2015. Microbial hotspots and hot moments in soil: concept & review. Soil Biology and Biochemistry, **83**:184–199.

LaMont, C. H. and P. A. Wiggins. 2015. Information-based inference in sloppy and singular models. arXiv preprint arXiv:1506.05855.

LaRowe, D. E. and J. P. Amend. 2015. Power limits for microbial life. Frontiers in Microbiology, **6**:718.

Lendenmann, U. and T. Egli. 1998. Kinetic models for the growth of Escherichia coli with mixtures of sugars under carbon-limited conditions. Biotechnology and Bioengineering, **59**:99–107.

Leon, M., M. L. Woods, A. J. Fedorec, and C. P. Barnes. 2016. A computational method for the investigation of multistable systems and its application to genetic switches. BMC Systems Biology, **10**:130.

Link, K., M. Stobb, J. Di Paola, K. Neeves, A. Fogelson, S. Sindi, and K. Leiderman. 2018. Link, K. G., et al. "A local and global sensitivity analysis of a mathematical model of coagulation and platelet deposition under flow. PloS One, **13**:e0200917.

Litchman, E., K. F. Edwards, and C. A. Klausmeier. 2015. Microbial resource utilization traits and trade-offs: implications for community structure, functioning, and biogeochemical impacts at present and in the future. Frontiers in Microbiology, **6**:254.

Liu, Y.-J., S.-J. Liu, H. L. Drake, and M. A. Horn. 2013. Consumers of 4-chloro-2-methylphenoxyacetic acid from agricultural soil and drilosphere harbor cadA, r/sdpA, and tfdA-like gene encoding oxygenases. FEMS Microbiology Ecology, **86**:114–129.

Lombardo, D. M. and W.-J. Rappel. 2017. Systematic reduction of a detailed atrial myocyte model. Chaos: An Interdisciplinary Journal of Nonlinear Science, **27**:093914.

López-Gutiérrez, J. C., S. Henry, S. Hallet, F. Martin-Laurent, G. Catroux, and L. Philippot. 2004. Quantification of a novel group of nitrate-reducing bacteria in the environment by real-time PCR. Journal of Microbiological Methods, **57**:399–407.

Luo, Y., A. Ahlström, S. D. Allison, N. H. Batjes, V. Brovkin, N. Carvalhais, A. Chappell, P. Ciais, E. A. Davidson, A. Finzi, et al. 2016. Toward more realistic projections of soil carbon dynamics by Earth system models. Global Biogeochemical Cycles, **30**:40–56.

Luo, Z., E. Wang, and O. J. Sun. 2017. Uncertain future soil carbon dynamics under global change predicted by models constrained by total carbon measurements. Ecological Applications, **27**:1001–1009.

Machta, B. B., R. Chachra, M. K. Transtrum, and J. P. Sethna. 2013. Parameter space compression underlies emergent theories and predictive models. Science, **342**:604–607.

Maiwald, T., H. Hass, B. Steiert, J. Vanlier, R. Engesser, A. Raue, F. Kipkeew, H. H. Bock, D. Kaschek, C. Kreutz, et al. 2016. Driving the model to its limit: profile likelihood based model reduction. PloS One, **11**:e0162366.

Manerkar, M. A., S. Seena, and F. Bärlocher. 2008. Q-RT-PCR for assessing archaea, bacteria, and fungi during leaf decomposition in a stream. Microbial Ecology, **56**:467–473.

Mannakee, B. K., A. P. Ragsdale, M. K. Transtrum, and R. N. Gutenkunst. 2016. Sloppiness and the geometry of parameter space. In Uncertainty in Biology, pages 271–299. Springer.

Manzoni, S., F. Moyano, T. Kätterer, and J. Schimel. 2016. Modeling coupled enzymatic and solute transport controls on decomposition in drying soils. Soil Biology and Biochemistry, **95**:275–287.

Manzoni, S. and A. Porporato. 2007. A theoretical analysis of nonlinearities and feedbacks in soil carbon and nitrogen cycles. Soil Biology and Biochemistry, **39**:1542–1556.

Manzoni, S. and A. Porporato. 2009. Soil carbon and nitrogen mineralization: theory and models across scales. Soil Biology and Biochemistry, **41**:1355–1379.

Manzoni, S., S. Schaeffer, G. Katul, A. Porporato, and J. Schimel. 2014. A theoretical analysis of microbial eco-physiological and diffusion limitations to carbon cycling in drying soils. Soil Biology and Biochemistry, **73**:69–83.

Martiny, J. B., S. E. Jones, J. T. Lennon, and A. C. Martiny. 2015. Microbiomes in light of traits: a phylogenetic perspective. Science, **350**:aac9323.

MATLAB. 2017. version 9.2 (R2017a). The MathWorks Inc., Natick, Massachusetts.

Mattingly, H. H., M. K. Transtrum, M. C. Abbott, and B. B. Machta. 2018. Maximizing the information learned from finite data selects a simple model. Proceedings of the National Academy of Sciences, **115**:1760–1765.

McCullagh, P. 2002. What is a statistical model? Annals of Statistics, pages 1225–1267.

McKay, M. D., R. J. Beckman, and W. J. Conover. 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, **21**:239–245.

Miltner, A., P. Bombach, B. Schmidt-Brücken, and M. Kästner. 2012. SOM genesis: microbial biomass as a significant source. Biogeochemistry, **111**:41–55.

Monga, O., P. Garnier, V. Pot, E. Coucheney, N. Nunan, W. Otten, and C. Chenu. 2014. Simulating microbial degradation of organic matter in a simple porous system using the 3-D diffusion-based model MOSAIC. Biogeosciences, **11**:2201–2209.

Monod, J. 1949. The growth of bacterial cultures. Annual Reviews in Microbiology, **3**:371–394.

Morris, M. D. 1991. Factorial sampling plans for preliminary computational experiments. Technometrics, **33**:161–174.

Mortensen, J. J., K. Kaasbjerg, S. L. Frederiksen, J. K. Nørskov, J. P. Sethna, and K. W. Jacobsen. 2005. Bayesian error estimation in density-functional theory. Physical Review Letters, **95**:216401.

Myers, C. R., R. N. Gutenkunst, and J. P. Sethna. 2007. Python unleashed on systems biology. Computing in Science & Engineering, **9**:34–37.

Nemenman, I. 2017. Renormalizing complex models: It is hard without Landau! Journal Club for Condensed Matter Physics. URL `https://www.condmatjclub.org/index.php?s=Nemenman`.

Nielsen, T. K., Z. Xu, E. Gözdereliler, J. Aamand, L. H. Hansen, and S. R. Sørensen. 2013. Novel insight into the genetic context of the cadAB genes from a 4-chloro-2-methylphenoxyacetic acid-degrading Sphingomonas. PLoS One, **8**:e83346.

Nikšić, T., M. Imbrišak, and D. Vretenar. 2017. Sloppy nuclear energy density functionals. II. Finite nuclei. Physical Review C, **95**:054304.

Nikšić, T. and D. Vretenar. 2016. Sloppy nuclear energy density functionals: Effective model reduction. Physical Review C, **94**:024333.

Nowak, K. M., A. Miltner, M. Gehre, A. Schaffer, and M. Kastner. 2010. Formation and fate of bound residues from microbial biomass during 2,4-D degradation in soil. Environmental Science & Technology, **45**:999–1006.

Pagel, H., J. Ingwersen, C. Poll, E. Kandeler, and T. Streck. 2014. Micro-scale modeling of pesticide degradation coupled to carbon turnover in the detritusphere: model description and sensitivity analysis. Biogeochemistry, **117**:185–204.

Pagel, H., C. Poll, J. Ingwersen, E. Kandeler, and T. Streck. 2016. Modeling coupled pesticide degradation and organic matter turnover: from gene abundance to process rates. Soil Biology and Biochemistry, **103**:349–364.

Parton, W., D. S. Schimel, C. Cole, and D. Ojima. 1987. Analysis of factors controlling soil organic matter levels in Great Plains Grasslands 1. Soil Science Society of America Journal, **51**:1173–1179.

Paul, E. A. 2014. Soil Microbiology, Ecology and Biochemistry. Academic press.

Pianosi, F., K. Beven, J. Freer, J. W. Hall, J. Rougier, D. B. Stephenson, and T. Wagener. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software, **79**:214–232.

Pianosi, F., F. Sarrazin, and T. Wagener. 2015. A Matlab toolbox for global sensitivity analysis. Environmental Modelling & Software, **70**:80–85.

Poeter, E. E., M. C. Hill, E. R. Banta, S. Mehl, and S. Christensen. 2006. UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation constructed using the JUPITER API. Technical report.

Poll, C., S. Marhan, J. Ingwersen, and E. Kandeler. 2008. Dynamics of litter carbon turnover and microbial abundance in a rye detritusphere. Soil Biology and Biochemistry, **40**:1306–1321.

Poll, C., H. Pagel, M. Devers-Lamrani, F. Martin-Laurent, J. Ingwersen, T. Streck, and E. Kandeler. 2010. Regulation of bacterial and fungal MCPA degradation at the soil–litter interface. Soil Biology and Biochemistry, **42**:1879–1887.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2007. Numerical recipes 3rd edition: The art of scientific computing. Cambridge University Press.

Raju, A., B. B. Machta, and J. P. Sethna. 2017. Information geometry and the renormalization group. arXiv preprint arXiv:1710.05787.

Rastetter, E. B. and J. J. Vallino. 2015. Ecosystems 80th and the Reemergence of Emergence. Ecosystems, **18**:735–739.

Raue, A., C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics, **25**:1923–1929.

Raue, A., M. Schilling, J. Bachmann, A. Matteson, M. Schelke, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, et al. 2013. Lessons learned from quantitative dynamical modeling in systems biology. PloS One, **8**:e74335.

Raupach, M., P. Rayner, D. Barrett, R. DeFries, M. Heimann, D. Ojima, S. Quegan, and C. Schmullius. 2005. Model–data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. Global Change Biology, **11**:378–397.

Revels, J., M. Lubin, and T. Papamarkou. 2016. Forward-mode automatic differentiation in Julia. arXiv preprint arXiv:1607.07892.

Ricci, M. and T. Levi-Civita. 1900. Méthodes de calcul différentiel absolu et leurs applications. Mathematische Annalen, **54**:125–201.

Risken, H. 1996. The Fokker-Planck Equation. In The Fokker-Planck Equation, pages 63–95. Springer.

Rosenbom, A. E., P. J. Binning, J. Aamand, A. Dechesne, B. F. Smets, and A. R. Johnsen. 2014. Does microbial centimeter-scale heterogeneity impact MCPA degradation in and leaching from a loamy agricultural soil? Science of the Total Environment, **472**:90–98.

Saleh, O., H. Pagel, E. Enowashu, M. Devers, F. Martin-Laurent, T. Streck, E. Kandeler, and C. Poll. 2016. Evidence for the importance of litter as a co-substrate for MCPA dissipation in an agricultural soil. Environmental Science and Pollution Research, **23**:4164–4175.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. 2008. Global sensitivity analysis: the primer. John Wiley & Sons.

Savenije, H. H. 2001. Equifinality, a blessing in disguise? Hydrological Processes, **15**:2835–2838.

Schaeffer, A., W. Amelung, H. Hollert, M. Kaestner, E. Kandeler, J. Kruse, A. Miltner, R. Ottermanns, H. Pagel, S. Peth, et al. 2016. The impact of chemical pollution on the resilience of soils under multiple stresses: a conceptual framework for future research. Science of the Total Environment, **568**:1076–1085.

Schimel, J. and S. M. Schaeffer. 2012. Microbial control over carbon cycling in soil. Frontiers in Microbiology, **3**:348.

Schöniger, A., W. A. Illman, T. Wöhling, and W. Nowak. 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection. Journal of Hydrology, **531**:96–110.

Schöniger, A., T. Wöhling, L. Samaniego, and W. Nowak. 2014. Model selection on solid ground: Rigorous comparison of nine ways to evaluate B ayesian model evidence. Water Resources Research, **50**:9484–9513.

Sierra, C., M. Müller, S. Trumbore, et al. 2012. Models of soil organic matter decomposition: the SoilR package, version 1.0. Geoscientific Model Development, **5**:1045–1060.

Sierra, C. A., S. Malghani, and M. Müller. 2015*a*. Model structure and parameter identification of soil organic matter models. Soil Biology and Biochemistry, **90**:197–203.

Sierra, C. A., S. E. Trumbore, E. A. Davidson, S. Vicca, and I. Janssens. 2015*b*. Sensitivity of decomposition rates of soil organic matter with respect to simultaneous changes in temperature and moisture. Journal of Advances in Modeling Earth Systems, **7**:335–356.

Sinsabaugh, R. L., S. Manzoni, D. L. Moorhead, and A. Richter. 2013. Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling. Ecology Letters, **16**:930–939.

Soetaert, K., T. Petzoldt, et al. 2010. Inverse modelling, sensitivity and monte carlo analysis in R using package FME. Journal of Statistical Software, **33**:1–28.

Stahr, K., E. Kandeler, L. Herrmann, and T. Streck. 2016. Bodenkunde und Standortlehre, volume 2967. UTB.

Standish, R. J., R. J. Hobbs, M. M. Mayfield, B. T. Bestelmeyer, K. N. Suding, L. L. Battaglia, V. Eviner, C. V. Hawkes, V. M. Temperton, V. A. Cramer, et al. 2014. Resilience in ecology: abstraction, distraction, or where the action is? Biological Conservation, **177**:43–51.

Stapor, P., D. Weindl, B. Ballnus, S. Hug, C. Loos, A. Fiedler, S. Krause, S. Hroß, F. Fröhlich, and J. Hasenauer. 2017. PESTO: Parameter EStimation TOolbox. Bioinformatics, **34**:705–707.

Stone, J. V. 2013. Bayes' rule: A tutorial introduction to Bayesian analysis. Sebtel Press.

Strigul, N., H. Dette, and V. B. Melas. 2009. A practical guide for optimal designs of experiments in the Monod model. Environmental Modelling & Software, **24**:1019–1026.

Strogatz, S. H. 2018. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press.

Tang, J. 2015. On the relationships between the Michaelis-Menten kinetics, reverse Michaelis-Menten kinetics, equilibrium chemistry approximation kinetics, and quadratic kinetics. Geoscientific Model Development, **8**:3823.

Tang, J. and W. Riley. 2013. A total quasi-steady-state formulation of substrate uptake kinetics in complex networks and an example application to microbial litter decomposition. Biogeosciences, **10**:8329–8351.

Tang, J. and Q. Zhuang. 2008. Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics. Journal of Geophysical Research: Biogeosciences, **113**.

Tierney, L. 1994. Markov chains for exploring posterior distributions. The Annals of Statistics, pages 1701–1728.

Todd-Brown, K., J. Randerson, W. Post, F. Hoffman, C. Tarnocai, E. Schuur, and S. Allison. 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences, **10**:1717–1736.

Todd-Brown, K. E., F. M. Hopkins, S. N. Kivlin, J. M. Talbot, and S. D. Allison. 2012. A framework for representing microbial decomposition in coupled climate models. Biogeochemistry, **109**:19–33.

Tönsing, C., J. Timmer, and C. Kreutz. 2014. Cause and cure of sloppiness in ordinary differential equation models. Physical Review E, **90**:023303.

Totsche, K. U., T. Rennert, M. H. Gerzabek, I. Kögel-Knabner, K. Smalla, M. Spiteller, and H.-J. Vogel. 2010. Biogeochemical interfaces in soil: the interdisciplinary challenge for soil science. Journal of Plant Nutrition and Soil Science, **173**:88–99.

Transtrum, M. K. 2016. Mbam source code. `https://github.com/mktranstrum/MBAM`.

Transtrum, M. K., G. Hart, and P. Qiu. 2014. Information topology identifies emergent model classes. arXiv preprint arXiv:1409.6203.

Transtrum, M. K., B. B. Machta, K. S. Brown, B. C. Daniels, C. R. Myers, and J. P. Sethna. 2015. Perspective: Sloppiness and emergent theories in physics, biology, and beyond. The Journal of Chemical Physics, **143**:07B201_1.

Transtrum, M. K., B. B. Machta, and J. P. Sethna. 2010. Why are nonlinear fits to data so challenging? Physical Review Letters, **104**:060201.

Transtrum, M. K., B. B. Machta, and J. P. Sethna. 2011. Geometry of nonlinear least squares with applications to sloppy models and optimization. Physical Review E, **83**:036701.

Transtrum, M. K. and P. Qiu. 2012. Optimal experiment selection for parameter estimation in biological differential equation models. BMC Bioinformatics, **13**:181.

Transtrum, M. K. and P. Qiu. 2014. Model reduction by manifold boundaries. Physical Review Letters, **113**:098701.

Transtrum, M. K. and P. Qiu. 2016. Bridging mechanistic and phenomenological models of complex biological systems. PLoS Computational Biology, **12**:e1004915.

Transtrum, M. K., A. T. Sarić, and A. M. Stanković. 2018. Information geometry approach to verification of dynamic models in power systems. IEEE Transactions on Power Systems, **33**:440–450.

van den Bos, A. and J. H. Swarte. 1993. Resolvability of the parameters of multiexponentials and other sum models. IEEE Transactions on Signal Processing, **41**:313.

Van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan. 2006. A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of Hydrology, **324**:10–23.

Van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang. 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Advances in Water Resources, **32**:1154–1169.

Verghese, G. 2009. Getting to the gray box: Some challenges for model reduction. In American Control Conference, 2009. ACC'09., pages 5–6. IEEE.

Villaverde, A. F. and J. R. Banga. 2014. Reverse engineering and identification in systems biology: strategies, perspectives and challenges. Journal of the Royal Society Interface, **11**:20130505.

Villaverde, A. F., A. Barreiro, and A. Papachristodoulou. 2016. Structural identifiability of dynamic systems biology models. PLoS Computational Biology, **12**:e1005153.

Vogel, L., V. Pot, D. Makowski, P. Garnier, and P. Baveye. 2018. To what extent do uncertainty and sensitivity analyses help unravel the influence of microscale physical and biological drivers in soil carbon dynamics models? Ecological Modelling, **383**:10–22.

Von Bertalanffy, L. 1969. General systems theory and psychiatryan overview. General Systems Theory, pages 33–46.

Vrugt, J. A. and B. A. Robinson. 2007. Improved evolutionary optimization from genetically adaptive multimethod search. Proceedings of the National Academy of Sciences, **104**:708–711.

Wang, G., S. Jagadamma, M. A. Mayes, C. W. Schadt, J. M. Steinweg, L. Gu, and W. M. Post. 2015. Microbial dormancy improves development and experimental validation of ecosystem model. The ISME Journal, **9**:226.

Wang, G., M. A. Mayes, L. Gu, and C. W. Schadt. 2014*a*. Representation of dormant and active microbial dynamics for ecosystem modeling. PLoS One, **9**:e89252.

Wang, G., W. M. Post, and M. A. Mayes. 2013. Development of microbial-enzyme-mediated decomposition model parameters through steady-state and dynamic analyses. Ecological Applications, **23**:255–272.

Wang, G., W. M. Post, M. A. Mayes, J. T. Frerichs, and J. Sindhu. 2012. Parameter estimation for models of ligninolytic and cellulolytic enzyme kinetics. Soil Biology and Biochemistry, **48**:28–38.

Wang, Y., B. Chen, W. R. Wieder, M. C. Leite, B. E. Medlyn, M. Rasmussen, M. J. Smith, F. B. Agusto, F. Hoffman, and Y. Luo. 2014*b*. Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. Biogeosciences, **11**:1817–1831.

Waterfall, J. J., F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna. 2006. Sloppy-model universality class and the Vandermonde matrix. Physical Review Letters, **97**:150601.

Weihermüller, L., A. Neuser, M. Herbst, and H. Vereecken. 2018. Problems associated to kinetic fitting of incubation data. Soil Biology and Biochemistry, **120**:260–271.

White, A., M. Tolman, H. D. Thames, H. R. Withers, K. A. Mason, and M. K. Transtrum. 2016. The limitations of model-based experimental design and parameter estimation in sloppy systems. PLoS Computational Biology, **12**:e1005227.

Wieder, W. R., S. D. Allison, E. A. Davidson, K. Georgiou, O. Hararuk, Y. He, F. Hopkins, Y. Luo, M. J. Smith, B. Sulman, et al. 2015. Explicitly representing soil microbial processes in Earth system models. Global Biogeochemical Cycles, **29**:1782–1800.

Wilson, K. G. 1971. Renormalization group and critical phenomena. I. Renormalization group and the Kadanoff scaling picture. Physical Review B, **4**:3174.

Wöhling, T., S. Gayler, E. Priesack, J. Ingwersen, H.-D. Wizemann, P. Högy, M. Cuntz, S. Attinger, V. Wulfmeyer, and T. Streck. 2013. Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and CLM3. 5. Water Resources Research, **49**:8200–8221.

Wutzler, T., S. A. Blagodatsky, E. Blagodatskaya, and Y. Kuzyakov. 2012. Soil microbial biomass and its activity estimated by kinetic respiration analysis–Statistical guidelines. Soil Biology and Biochemistry, **45**:102–112.

Wutzler, T. and M. Reichstein. 2013. Priming and substrate quality interactions in soil organic matter models. Biogeosciences, **10**:2089–2103.

Zamora-Sillero, E., M. Hafner, A. Ibig, J. Stelling, and A. Wagner. 2011. Efficient characterization of high-dimensional parameter spaces for systems biology. BMC Systems Biology, **5**:142.

Zelenev, V., A. Van Bruggen, and A. Semenov. 2005. Modeling wave-like dynamics of oligotrophic and copiotrophic bacteria along wheat roots in response to nutrient input from a growing root tip. Ecological Modelling, **188**:404–417.

Zi, Z. 2011. Sensitivity analysis approaches applied to systems biology models. IET Systems Biology, **5**:336–346.

Appendix

Model Definitions, Initial Conditions and Parameter Values

Table A1 Model parameter symbols, descriptions, base values of parameters of the minimal microbial soil carbon model (M=2, N=4; German et al. 2012, Table 3).

Symbol	Description	Value	Unit
$\mathrm{V_{max}}$	Maximum cycling rate of soil carbon	0.0019	$\left[\frac{1}{h\ cm^3\ mg}\right]$
$\mathrm{K_S}$	Half-saturation constant	1.24	$\left[\frac{mg}{cm^3}\right]$
$\mathrm{k_B}$	First-order cycling rate for microbial biomass	0.0005	$\left[\frac{1}{h}\right]$
Y	Microbial carbon use efficiency	0.134	$[-]$
I	External carbon input	0.001	$\left[\frac{mg}{cm^3\ h}\right]$
$\mathrm{C_B}(0)$	Initial microbial biomass carbon	2.0	$\left[\frac{mg}{cm^3}\right]$
$\mathrm{C_S}(0)$	Initial soil organic carbon	100.0	$\left[\frac{mg}{cm^3}\right]$

Table A2 Carbon stocks and governing differential equations of the extended NICA model (M=10, N=15).

C stock	Differential equation	
i-s microbial biomass $[\frac{mg\,C}{g}]$	$\frac{dC_{b,is}}{dt} = r_{is}C_{b,is}\left(\mu(C_{s,is}) - a(C_{s,is})\right)$	(1)
l-s microbial biomass $[\frac{mg\,C}{g}]$	$\frac{dC_{b,ls}}{dt} = r_{ls}C_{b,ls}\left(\mu(C_{s,ls}) - a(C_{s,ls})\right)$	(2)
Physiological state index of i-s decomposer $[-]$	$\frac{dr_{is}}{dt} = \mu(C_{s,is}) \cdot (\Phi(C_{s,is}) - r_{is})$	(3)
Physiological state index of l-s decomposer $[-]$	$\frac{dr_{ls}}{dt} = \mu(C_{s,ls}) \cdot (\Phi(C_{s,ls}) - r_{ls})$	(4)
i-s dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{s,is}}{dt} = r_{is}C_{b,is}\left(\frac{1}{Y_s}\mu(C_{s,is}) - m(C_{s,is})\right) + I_{is}$	(5)
l-s dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{s,ls}}{dt} = r_{ls}C_{b,ls}\left(\frac{1}{Y_s}\mu(C_{s,ls}) - q(C_h)Y_r\right) + I_{ls}$	(6)
Insoluble soil organic matter $[\frac{mg\,C}{g}]$	$\frac{dC_h}{dt} = r_{is}C_{b,is}a(C_{s,is}) + r_{ls}C_{b,ls}\left(a(C_{s,ls}) - q(C_h)\right)$	(7)

Table A3 Biokinetic functions and composite parameter expressions of the extended NICA model (M=10, N=15).

Description	Expression	Unit	
Specific rate of initial-stage decomposer growth	$\mu(C_{s,is}) = \frac{\mu_{max,is} C_{s,is}}{C_{s,is}+K_{s,is}}$	$\left[\frac{1}{d}\right]$	(8)
Specific rate of late-stage decomposer growth	$\mu(C_{s,ls}) = \frac{\mu_{max,ls} C_{s,ls}}{C_{s,ls}+K_{s,ls}}$	$\left[\frac{1}{d}\right]$	(9)
Specific rate of organic matter decomposition	$q(C_h) = \frac{q_{max} C_h}{C_h+K_h}$	$\left[\frac{1}{d}\right]$	(10)
Substrate dependent specific death rate of initial-stage decomposer	$a(C_{s,is}) = \frac{a_{max}}{1+K_{a,is} C_{s,is}}$	$\left[\frac{1}{d}\right]$	(11)
Substrate dependent specific death rate of late-stage decomposer	$a(C_{s,ls}) = \frac{a_{max}}{1+K_{a,ls} C_{s,ls}}$	$\left[\frac{1}{d}\right]$	(12)
Specific rate of maintenance respiration of initial-stage decomposer	$m(C_{s,is}) = \frac{m_{max} C_{s,is}}{K_m+C_{s,is}}$	$\left[\frac{1}{d}\right]$	(13)
Limiting factor for activity increase of initial-stage decomposer	$\Phi(C_{s,is}) = \frac{C_{s,is}}{k_{rC,is}+C_{s,is}}$	$[-]$	(14)
Limiting factor for activity increase of late-stage decomposer	$\Phi(C_{s,ls}) = \frac{C_{s,ls}}{k_{rC,ls}+C_{s,ls}}$	$[-]$	(15)

Table A4 Model parameter symbols, descriptions, base values of parameters of the extended NICA model (M=10, N=15; Ingwersen et al. 2008, Table 2).

Symbol	Description	Value	Unit
$\mu_{\mathrm{max,is}}$	Maximal specific growth rate of initial-stage decomposer	25.5	$\left[\frac{1}{d}\right]$
$\mu_{\mathrm{max,ls}}$	Maximal specific growth rate of late-stage decomposer	2.59	$\left[\frac{1}{d}\right]$
$\mathrm{q_{max}}$	Maximal specific rate of organic matter decomposition	1.62	$\left[\frac{1}{d}\right]$
$\mathrm{a_{max}}$	Maximal specific death/reutilization rate of decomposer	1.309	$\left[\frac{1}{d}\right]$
$\mathrm{m_{max}}$	Maximal specific maintenance rate of initial-stage decomposer	0.25	$\left[\frac{1}{d}\right]$
$\mathrm{K_{s,is}}$	MichaelisMenten constant for initial-stage decomposer growth	0.264	$\left[\frac{mg\,C}{g}\right]$

$K_{s,ls}$	MichaelisMenten constant for late-stage decomposer growth	0.264	$\left[\frac{mg\,C}{g}\right]$
K_h	MichaelisMenten constant for organic matter decomposition	13.75	$\left[\frac{mg\,C}{g}\right]$
K_m	MichaelisMenten constant for maintenance respiration of initial-stage decomposer	0.001	$\left[\frac{mg\,C}{g}\right]$
$k_{rC,is}$	Inhibition constant for C-dependent initial-stage decomposer activity	1.3	$\left[\frac{mg\,C}{g}\right]$
$k_{rC,ls}$	Inhibition constant for C-dependent late-stage decomposer activity	1.3	$\left[\frac{mg\,C}{g}\right]$
$K_{a,is}$	Inhibition constant for initial-stage decomposer death rate	12.425	$\left[\frac{g}{mg\,C}\right]$
$K_{a,ls}$	Inhibition constant for late-stage decomposer death rate	12.425	$\left[\frac{g}{mg\,C}\right]$
Y_s	Efficiency of substrate uptake	0.848	$[-]$
Y_r	Efficiency of organic matter mineralisation and biomass reutilisation	0.50	$[-]$
I_{is}	i-s litter carbon input	0.001	$\left[\frac{mg\,C}{g\,d}\right]$
I_{ls}	l-s litter carbon input	0.001	$\left[\frac{mg\,C}{g\,d}\right]$

Table A5 Carbon stocks and governing differential equations of the MEND model (M=10, N=19).

C stock	Differential equation	
Lignocellulose-like particulate organic carbon $[\frac{mg\,C}{g}]$	$\frac{dP_1}{dt} = I_{P_1} + (1-g_D)\cdot F_{12} - F_1$	(1)
Starch-like particulate organic carbon $[\frac{mg\,C}{g}]$	$\frac{dP_2}{dt} = I_{P_2} - F_2$	(2)
Mineral-associated organic carbon $[\frac{mg\,C}{g}]$	$\frac{dM}{dt} = (1-f_D)\cdot(F_1+F_2) - F_3$	(3)
Dissolved organic carbon (DOC) $[\frac{mg\,C}{g}]$	$\frac{dD}{dt} = I_D + f_D\cdot(F_1+F_2) + g_D F_{12} + F_3 - F_6 + \left(F_{14,EP_1} + F_{14,EP_2}F_{14,EM}\right) - (F_4 - F_5)$	(4)
Adsorbed phase of DOC $[\frac{mg\,C}{g}]$	$\frac{dQ}{dt} = F_4 - F_5$	(5)
Active microbial biomass $[\frac{mg\,C}{g}]$	$\frac{dBA}{dt} = F_6 - (F_7 - F_8) - (F_9 - F_{10}) - F_{12} - \left(F_{14,EP_1} + F_{14,EP_2}F_{14,EM}\right)$	(6)
Dormant microbial biomass $[\frac{mg\,C}{g}]$	$\frac{dBD}{dt} = (F_7 - F_8) - F_{11}$	(7)

P1 degraded enzymes $[\frac{mg\,C}{g}]$	$\frac{dEP_1}{dt} = F_{13,EP_1} - F_{14,EP_1}$	(8)
P2 degraded enzymes $[\frac{mg\,C}{g}]$	$\frac{dEP_2}{dt} = F_{13,EP_2} - F_{14,EP_2}$	(9)
M degraded enzymes $[\frac{mg\,C}{g}]$	$\frac{dEM}{dt} = F_{13,EM} - F_{14,EM}$	(10)
CO_2 $[\frac{mg\,C}{g}]$	$\frac{dCO_2}{dt} = (F_9 + F_{10}) + F_{11}$	(11)

Table A6 Biokinetic functions and composite parameter expressions of the MEND model (M=10, N=19).

Description	Expression	
P1 decomposition	$F_1 = \frac{V_{P1} \cdot EP_1 \cdot P_1}{K_{P_1} + P_1}$	(12)
P2 decomposition	$F_2 = \frac{V_{P2} \cdot EP_2 \cdot P_2}{K_{P_2} + P_2}$	(13)
Mineral-associated organic carbon decomposition	$F_3 = \frac{V_M \cdot EM \cdot M}{K_M + M}$	(14)
Adsorption of DOC	$F_4 = K_{ads} \cdot (1 - Q/Q_{max}) \cdot D$	(15)
Desorption of DOC	$F_5 = K_{des} \cdot (Q/Q_{max})$	(16)
DOC uptake by microbes	$F_6 = \frac{1}{Y_G} (V_D + m_R) \frac{D \cdot BA}{K_D + D}$	(17)
Dormancy flux	$F_7 = \left(1 - \frac{D}{K_D + D}\right) \cdot m_R \cdot BA$	(18)
Reactivation flux	$F_8 = \frac{D}{K_D + D} \cdot m_R \cdot BD$	(19)
BA growth respiration	$F_9 = \left(\frac{1}{Y_G} - 1\right) \frac{V_D \cdot BA \cdot D}{K_D + D}$	(20)
BA maintenance respiration	$F_{10} = \left(\frac{1}{Y_G} - 1\right) \frac{m_R \cdot BA \cdot D}{K_D + D}$	(21)
BD maintenance respiration	$F_{11} = \beta \cdot m_R \cdot BD$	(22)
BA mortality	$F_{12} = (1 - p_{EP} - p_{EM}) \cdot m_R \cdot BA$	(23)
Synthesis of enzymes for P1	$F_{13,EP_1} = \frac{P_1}{P_1 + P_2} \cdot p_{EP} \cdot m_R \cdot BA$	(24)
Synthesis of enzymes for P2	$F_{13,EP_2} = \frac{P_2}{P_1 + P_2} \cdot p_{EP} \cdot m_R \cdot BA$	(25)
Synthesis of enzymes for M	$F_{13,EM} = p_{EM} \cdot m_R \cdot BA$	(26)
Turnover of enzymes	$F_{14,EP_1} = r_E \cdot EP_1$ $F_{14,EP_2} = r_E \cdot EP_2$ $F_{14,EM} = r_E \cdot EM$	(27)

Table A7 Model parameter symbols, descriptions, base values of parameters of the MEND model (M=10, N=19; Wang et al. 2015, Figure S2, Gelisol).

Symbol	Description	Value	Unit
V_{P1}	Maximum specific decomposition rate for P1	1.6	$\left[\frac{1}{h}\right]$
V_{P2}	Maximum specific decomposition rate for P2	38.0	$\left[\frac{1}{h}\right]$
K_{P1}	Half-saturation constant for P1 decomposition	50.0	$\left[\frac{mg\,C}{g}\right]$
K_{P2}	Half-saturation constant for P2 decomposition	18.0	$\left[\frac{mg\,C}{g}\right]$
V_M	Maximum specific decomposition rate for M	1.1	$\left[\frac{1}{h}\right]$
K_M	Half-saturation constant for M decomposition	455.0	$\left[\frac{mg\,C}{g}\right]$
V_D	Maximum specific uptake rate of D for growth	0.04	$\left[\frac{1}{h}\right]$
K_M	Half-saturation constant for uptake of D	0.19	$\left[\frac{mg\,C}{g}\right]$
m_R	Specific maintenance rate of BA	0.033	$\left[\frac{1}{h}\right]$
β	Ratio of dormant maintenance rate to m_R	0.001	$[-]$
Y_G	True growth yield	0.27	$[-]$
f_D	Fraction of decomposed P1 and P2 allocated to D	0.7	$[-]$
g_D	Fraction of dead BA allocated to D	0.3	$[-]$
p_{EP}	Fraction of m_R for production of EP1 and EP2	0.05	$[-]$
p_{EM}	Fraction of m_R for production of EM	0.05	$[-]$
r_E	Turnover rate of EP1, EP2, and EM	0.0025	$\left[\frac{1}{h}\right]$
Q_{max}	Maximum DOC sorption capacity	3.5	$\left[\frac{mg\,C}{g}\right]$
K_{des}	Desorption rate	0.048	$\left[\frac{mg\,C}{g\,h}\right]$
K_{ads}	Adsorption rate	0.48	$\left[\frac{1}{h}\right]$

Table A8 Carbon stocks and governing differential equations of the trait-based microbial soil carbon model (M=7, N=24).

C stock	Differential equation	
Stable soil organic C substrates $[\frac{g\,C}{m^3}]$	$\frac{dC_S}{dt} = I_L - D$	(1)
Soluble organic C $[\frac{g\,C}{m^3}]$	$\frac{dC_D}{dt} = D + M_B\,(1+\gamma) + M_{B,D}\,(1+\gamma) + E_D + P_{D\to A}\gamma - U - L_D$	(2)
Enzymatic C $[\frac{g\,C}{m^3}]$	$\frac{dC_E}{dt} = E_P - E_D - L_E$	(3)
C in active microbial biomass $[\frac{g\,C}{m^3}]$	$\frac{dC_B}{dt} = \frac{eU - R_M - E_P + P_{D\to A}}{1+\gamma} - P_{A\to D} - M_B$	(4)
C in dormant microbial biomass $[\frac{g\,C}{m^3}]$	$\frac{dC_{B,D}}{dt} = P_{A\to D} - P_{D\to A} - M_{B,D}$	(5)

Table A9 Biokinetic functions and composite parameter expressions of the trait-based microbial soil carbon model (M=7, N=24).

Description	Expression	Unit	
Microbial uptake	$U = h_D(s)C_D$	$\left[\frac{gC}{m^3\,d}\right]$	(6)
Mortality of active microbial biomass	$M_B = k_B C_B$	$\left[\frac{gC}{m^3\,d}\right]$	(7)
Mortality of dormant microbial biomass	$M_{B,D} = k_B/10 \cdot C_{B,D}$	$\left[\frac{gC}{m^3\,d}\right]$	(8)
Enzyme decay rate	$E_D = k_E \cdot C_E$	$\left[\frac{gC}{m^3\,d}\right]$	(9)
Enzyme production rate	$E_P = h_E(s)\left(C_{E,0} - C_E\right)$	$\left[\frac{gC}{m^3\,d}\right]$	(10)
Transfer from dormant to active population	$P_{D\to A} = k_{D\to A} f_{D\to A}(\psi) C_{B,D}$	$\left[\frac{gC}{m^3\,d}\right]$	(11)
Transfer from active to dormant population	$P_{A\to D} = k_{A\to D} f_{A\to D}(\psi) C_B$	$\left[\frac{gC}{m^3\,d}\right]$	(12)
Maintenance respiration	$R_M = k_M C_B$	$\left[\frac{gC}{m^3\,d}\right]$	(13)
Leaching of dissolved organic C	$L_D = C_D L_s Z_r^{-1} \cdot (\rho_b K_d + n \cdot s)^{-1}$	$\left[\frac{gC}{m^3\,d}\right]$	(14)
Leaching of enzymes	$L_E = L_s C_E Z_r^{-1} \cdot (\rho_b K_d + n \cdot s)^{-1}$	$\left[\frac{gC}{m^3\,d}\right]$	(15)
Transfer coefficient for dissolved organic C	$h_D(s) = \delta^{-2} D_D(s) \cdot (\rho_b K_d + n \cdot s)^{-1} \nu$	$\left[\frac{1}{d}\right]$	(16)
Diffusivity of dissolved organic C in bulk soil	$D_D(s) = ((s_{th} - s) \cdot (-1 + s_{th})^{-1})^{m2} \cdot D_{D,0} \cdot n^{m1} \cdot (1 - s_{th})^{m1}$	$\left[\frac{m^2}{s}\right]$	(17)
Transfer coefficient for enzymes	$h_E(s) = \delta^{-2} D_E(s) \cdot (\rho_b K_d + n \cdot s)^{-1} \nu$	$\left[\frac{1}{d}\right]$	(18)

Diffusivity of enzymes in bulk soil	$D_D(s) = ((s_{th} - s) \cdot (-1 + s_{th})^{-1})^{m2} \cdot D_{E,0} \cdot n^{m1} \cdot (1 - s_{th})^{m1}$	$\left[\frac{m^2}{s}\right]$	(19)
Switching function for active-dormant state transition	$f_{A\to D}(\psi) = \frac{(-\psi)^w}{(-\psi)^w + (-\psi_{A\to D})^w}$	$[-]$	(20)
Switching function for dormant-active state transition	$f_{D\to A}(\psi) = \frac{(-\psi_{D\to A})^w}{(-\psi)^w + (-\psi_{D\to A})^w}$	$[-]$	(21)
Soil matric potential	$\psi = s^{-b} \cdot \psi_{sat}$	$[MPa]$	(22)

Table A10 Model parameter symbols, descriptions, base values of parameters of the trait-based microbial soil carbon model (M=7, N=24; Manzoni et al. 2014, Table 2).

Symbol	Description	Value	Unit
b	Exponent of the water retention curve	4.9	$[-]$
$\mathrm{C_{E,0}}$	Enzyme concentration outside the microbial cell	1.0	$\left[\frac{gC}{m^3}\right]$
$\mathrm{D_{D,0}}$	Diffusivity of dissolved organic C in pure water	$8.1e^{-10}$	$\left[\frac{m^2}{s}\right]$
δ	Characteristic distance between microbial cells and substrate	$1e^{-4}$	$[m]$
e	Growth efficiency	0.5	$[-]$
γ	Fixed ratio for constitutive osmolyte production	0.026	$[-]$
$\mathrm{I_L}$	Litter carbon input (fixed)	0.9	$\left[\frac{gC}{m^3\,d}\right]$
$\mathrm{k_{A\to D}}$	Maximum rate of transition from active to dormant state	1.0	$\left[\frac{1}{d}\right]$
$\mathrm{k_B}$	Mortality rate of active population	0.012	$\left[\frac{1}{d}\right]$
$\mathrm{K_d}$	Solid-liquid partition coefficient	$1e-5$	$\left[\frac{m^3}{g}\right]$
$\mathrm{k_D}$	Maximum rate of decomposition	$1e-3$	$\left[\frac{1}{d}\right]$
$\mathrm{k_{D\to A}}$	Maximum rate of transition from dormant to active state	$k_{A\to D}$	$\left[\frac{1}{d}\right]$
$\mathrm{k_E}$	Enzyme de-activation rate	$5e-4$	$\left[\frac{1}{d}\right]$
$\mathrm{k_M}$	Maintenance respiration rate	0.022	$\left[\frac{1}{d}\right]$
$\mathrm{m_1}$	Empirical exponent	1.5	$[-]$
$\mathrm{m_2}$	Empirical exponent	2.5	$[-]$
n	Soil porosity	0.43	$[-]$
ν	Scaling coefficient	6.0	$[-]$
$\psi_{\mathrm{A\to D}}$	Water potential at 50% of the maximum rate $k_{A\to D}$	0.4	$[MPa]$

ψ_{sat}	Soil water potential at saturation	-0.002	$[MPa]$
ρ_{b}	Soil bulk density	$1.2e^{6}$	$\left[\frac{g}{m^3}\right]$
s	Soil moisture	0.6	$[-]$
$\mathrm{s_{th}}$	Diffusion threshold	0.18	$[-]$
w	Sensitivity parameter for the switching functions	4.0	$[-]$
$\mathrm{Z_r}$	Soil depth	0.4	$[m]$

Table A11 Carbon stocks and governing differential equations of the full PECCAD ODE model (M=12, N=59).

C stock	Differential equation	
Bacteria $[\frac{mg\,C}{g}]$	$\frac{dC_B}{dt} = r_B C_B \left(\mu_{B,hiq} + \mu_{B,loq} - a_B\right)$	(1)
Fungi $[\frac{mg\,C}{g}]$	$\frac{dC_F}{dt} = r_F C_F \left(\mu_{F,hiq} + \mu_{F,loq} - a_F - a_{max-F}(1 - Y_{r-F}) - m_{F,P}\right)$	(2)
Specific pesticide degraders $[\frac{mg\,C}{g}]$	$\frac{dC_{BP}}{dt} = r_{BP} C_{BP} \left(\mu_{BP,P} + \mu_{BP,hiq} + \mu_{BP,loq} - a_{BP}\right)$	(3)
Physiological state index of bacteria $[-]$	$\frac{dr_B}{dt} = (\mu_{B,hiq} + \mu_{B,loq}) \cdot (\Phi_B - r_B)$	(4)
Physiological state index of fungi $[-]$	$\frac{dr_F}{dt} = (\mu_{F,hiq} + \mu_{F,loq}) \cdot (\Phi_F - r_F)$	(5)
Physiological state index of specific pesticide degraders $[-]$	$\frac{dr_{BP}}{dt} = (\mu_{BP,P} + \mu_{BP,hiq} + \mu_{BP,loq}) \cdot (\Phi_{BP} - r_{BP})$	(6)
hiq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{hiq}}{dt} = I_{hiq}(t) + r_B C_B \left(-\frac{1}{Y_{s-B,hiq}}\mu_{B,hiq} - m_{B,hiq}\right) + r_F C_F \left(\frac{1}{Y_{s-F,hiq}}\mu_{F,hiq} + Y_{R-F,P} q_{F,P}\right) + r_{BP} C_{BP} \left(\frac{1}{Y_{s-BP,hiq}}\mu_{BP,hiq} - m_{BP,hiq}\right)$	(7)
Sorbed phase	$C_{hiq-s} = \frac{\rho_B}{\theta} K_{d-hiq} C_{hiq}$	(8)
loq dissolved organic C $[\frac{mg\,C}{g}]$	$\frac{dC_{loq}}{dt} = I_{loq}(t) + r_B C_B \left(-\frac{1}{Y_{s-B,loq}}\mu_{B,loq} - m_{B,loq} + q_B Y_{r-B}\right) + r_F C_F \left(-\frac{1}{Y_{s-F,loq}}\mu_{F,loq} + q_F Y_{r-F}\right) + r_{BP} C_{BP} \left(-\frac{1}{Y_{s-BP,loq}}\mu_{BP,loq} - m_{BP,loq} + q_B Y_{r-B}\right)$	(9)
Sorbed phase	$C_{loq-s} = \frac{\rho_B}{\theta} K_{d-loq} C_{loq}$	(10)

Insoluble soil organic matter C $[\frac{mg\,C}{g}]$	$\frac{dC_I}{dt} = r_B C_B\,(a_B - q_B) + r_F C_F\,(a_F - q_F) + r_{BP} C_{BP}\,(a_{BP} - q_B)$	(11)
Pesticide C $[\frac{mg\,C}{g}]$	$\frac{d(C_P + C_{P-s})}{dt} = r_{BP} C_{BP}\left(-\frac{1}{Y_{s-BP,P}}\mu_{BP,P} - m_{BP,P}\right) - r_F C_F q_{F,P}$	(12)
Sorbed phase	$C_{P-s} = K_{F-P}\left(C_P \frac{1000\rho_B M_P}{9\theta M_c}\right)^{n_{F-P}}$	(13)
CO_2 $[\frac{mg\,C}{g}]$	$\frac{dCO_2}{dt} = r_B C_B\left(\frac{1 - Y_{s-B,hiq}}{Y_{s-B,hiq}}\mu_{B,hiq} + \frac{1 - Y_{s-B,loq}}{Y_{s-B,loq}}\mu_{B,loq}\right)$ $+ r_B C_B\left(m_{B,hiq} + m_{B,loq} + q_B(1 - Y_{r-B})\right)$ $+ r_{BP} C_{BP}\left(\frac{1 - Y_{s-BP,hiq}}{Y_{s-BP,hiq}}\mu_{BP,hiq} + \frac{1 - Y_{s-BP,loq}}{Y_{s-BP,loq}}\mu_{BP,loq}\right)$ $+ r_{BP} C_{BP}\left(\frac{1 - Y_{s-BP,P}}{Y_{s-BP,P}}\mu_{BP,P} + m_{BP,hiq} + m_{BP,loq} + m_{BP,P}\right)$ $+ r_{BP} C_{BP} q_B(1 - Y_{r-B})$ $+ r_F C_F\left(\frac{1 - Y_{s-F,hiq}}{Y_{s-F,hiq}}\mu_{F,hiq} + \frac{1 - Y_{s-F,loq}}{Y_{s-F,loq}}\mu_{F,loq} + q_F(1 - Y_{r-F})\right)$ $+ r_F C_F\left(a_{max-F}(1 - Y_{r-F}) + m_{F,P} + q_{F,P}(1 - Y_{R-F,P})\right)$	(14)
Litter C $[mg\,C]$	$\frac{dC_{L,tot}}{dt} = -k_L C_{L,tot}$	(15)
Litter derived CO_2 $[mg\,C]$	$\frac{dCO_{2-L}}{dt} = (1 - Y_{L,hiq} \cdot f_L - Y_{L,loq} \cdot (1 - f_L)) \cdot k_L \cdot C_{L,tot}$	(16)

Table A12 Biokinetic functions and composite parameter expressions of the full PECCAD ODE model (M=12, N=59).

Description	Expression	Unit	
Rate of total litter decomposition	$k_L = c_L + \left(\frac{t}{t^2+b_L}\right)^3$	$\left[\frac{1}{d}\right]$	(17)
Fraction of hiq litter on total decomposed litter	$f_L = \frac{k_l - c_L}{k_{L,max}}$	$[-]$	(18)
Maximum rate of total litter decomposition	$k_{L,max} = c_L + \left(\frac{1}{2\cdot\sqrt{b_L}}\right)^3$	$\left[\frac{1}{d}\right]$	(19)
hiq litter derived DOC	$I_{hiq}(t) = Y_{L,hiq} \cdot f_L \cdot k_L \cdot C_{L,tot}$	$\left[\frac{mg\,C}{d}\right]$	(20)
loq litter derived DOC	$I_{loq}(t) = Y_{L,loq} \cdot (1 - f_L) \cdot k_L \cdot C_{L,tot}$	$\left[\frac{mg\,C}{d}\right]$	(21)
	with $i = \{hiq, loq\}$:		
Substrate dependent specific growth rate of bacteria	$\mu_{B,i} = \frac{\mu_{max-B} k_{B,i} C_i}{\mu_{max-B} + \sum_i k_{B,i} C_i}$	$\left[\frac{1}{d}\right]$	(22)
Substrate dependent specific growth rate of fungi	$\mu_{F,i} = \frac{\mu_{max-F} k_{F,i} C_i}{\mu_{max-F} + \sum_i k_{F,i} C_i}$	$\left[\frac{1}{d}\right]$	(23)
Substrate dependent specific rate of maintenance respiration of bacteria	$m_{B,i} = \frac{m_{max-B} k_{m-B,i} C_i}{m_{max-B} + \sum_i k_{m-B,i} C_i}$	$\left[\frac{1}{d}\right]$	(24)
	with $i = \{P, hiq, loq\}$:		
Substrate dependent specific growth rate of bacterial pesticide degraders	$\mu_{BP,i} = \frac{\mu_{max-BP} k_{BP,i} C_i}{\mu_{max-BP} + \sum_i k_{BP,i} C_i}$	$\left[\frac{1}{d}\right]$	(25)
Substrate dependent specific rate of maintenance respiration of bacterial pesticide degraders	$m_{BP,i} = \frac{m_{max-BP} k_{m-BP,i} C_i}{m_{max-BP} + \sum_i k_{m-BP,i} C_i}$	$\left[\frac{1}{d}\right]$	(26)

Co-metabolic pesticide consumption rate of fungi	$q_{F,P} = \left(T_{y-F} \cdot \left(\mu_{F,hiq} + \mu_{F,loq}\right) + k_{F,P}\right) \cdot \frac{C_P}{K_{s-F,P}+C_P}$	$\left[\frac{1}{d}\right]$	(27)
Endogenous maintenance rate of fungi due to co-metabolic pesticide degradation	$m_{F,P} = \frac{q_{F,P}}{T_{F,P}}$	$\left[\frac{1}{d}\right]$	(28)
Specific death rate of bacteria	$a_B = \frac{a_{max-B}}{1+K_{a-B,hiq}C_{hiq}+K_{a-B,loq}C_{loq}}$	$\left[\frac{1}{d}\right]$	(29)
Specific death rate of fungi	$a_F = \frac{a_{max-F}}{1+K_{a-F,hiq}C_{hiq}+K_{a-F,loq}C_{loq}}$	$\left[\frac{1}{d}\right]$	(30)
Specific death rate of bacterial pesticide degraders	$a_{BP} = \frac{a_{max-BP}}{1+K_{a-BP,P}C_P+K_{a-BP,hiq}C_{hiq}+K_{a-BP,loq}C_{loq}}$	$\left[\frac{1}{d}\right]$	(31)
Specific rate of insoluble SOM decomposition by bacteria and bacterial pesticide degraders	$q_B = \frac{q_{max-B}C_I}{K_{I-B}+C_I}$	$\left[\frac{1}{d}\right]$	(32)
Specific rate of insoluble SOM decomposition by fungi	$q_F = \frac{q_{max-F}C_I}{K_{I-F}+C_I}$	$\left[\frac{1}{d}\right]$	(33)
Limiting factor of activity increase of bacteria	$\Phi_B = \frac{C_{hiq}/k_{r-B,hiq}+C_{loq}/k_{r-B,loq}}{1+C_{hiq}/k_{r-B,hiq}+C_{loq}/k_{r-B,loq}}$	$[-]$	(34)
Limiting factor of activity increase of fungi	$\Phi_F = \frac{C_{hiq}/k_{r-F,hiq}+C_{loq}/k_{r-F,loq}}{1+C_{hiq}/k_{r-F,hiq}+C_{loq}/k_{r-F,loq}}$	$[-]$	(35)
Limiting factor of activity increase of bacterial pesticide degraders	$\Phi_{BP} = \frac{C_P/k_{r-BP,P}+C_{hiq}/k_{r-BP,hiq}+C_{loq}/k_{r-BP,loq}}{1+C_P/k_{r-BP,P}+C_{hiq}/k_{r-BP,hiq}+C_{loq}/k_{r-BP,loq}}$	$[-]$	(36)

Table A13 Model parameter symbols, descriptions, values of optimal parameters of the full PECCAD ODE model (M=12, N=59) calibrated on the data of the MCPA + Litter treatment, 95% highest posterior density interval (HDI) and units.

Symbol	Description	MCPA + Litter	HDI	Unit
a_{max-B}	Maximal specific death rate of bacteria	1.21	$[0.004, 7.89]$	$\left[\frac{1}{d}\right]$
a_{max-BP}	Maximal specific death rate of bacterial pesticide degraders	0.15	$[0.004, 3.79]$	$\left[\frac{1}{d}\right]$
a_{max-F}	Maximal specific death rate of fungi	10.88	$[4.70, 23.94]$	$\left[\frac{1}{d}\right]$
$K_{a-B,hiq}$	Inhibition coefficient of bacterial death rate in response to hiq DOC	9.74	$[0.023, 622.48]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-B,loq}$	Inhibition coefficient of bacterial death rate in response to loq DOC	16.68	$[0.39, 537.70]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,hiq}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to hiq DOC	123.96	$[0.235, 2.64e3]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,loq}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to loq DOC	29.12	$[0.093, 206.64]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,P}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to pesticide	19.41	$[0.147, 441.90]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-F,hiq}$	Inhibition coefficient of fungal death rate in response to hiq DOC	161.16	$[22.94, 1.27e3]$	$\left[\frac{g}{mg\,C}\right]$
$K_{a-F,loq}$	Inhibition coefficient of fungal death rate in response to loq DOC	15.92	$[5.73, 90.0]$	$\left[\frac{g}{mg\,C}\right]$
$k_{B,hiq}$	hiq DOC growth substrate affinity coefficient of bacteria	231.20	$[20.02, 1.95e3]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{B,loq}$	loq DOC growth substrate affinity coefficient of bacteria	5.83	$[0.18, 25.43]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,hiq}$	hiq DOC growth substrate affinity coefficient of bacterial pesticide degraders	513.89	$[1.54, 4.38e3]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,loq}$	loq DOC growth substrate affinity coefficient of bacterial pesticide degraders	335.55	$[12.25, 1.07e3]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,P}$	pesticide growth substrate affinity coefficient of bacterial pesticide degraders	461.64	$[1.43, 3.52e^4]$	$\left[\frac{g}{mg\,C\,d}\right]$

K_{d-hiq}	Linear sorption coefficient of hiq DOC	1.13	$[0.115, 3.574]$	$\left[\frac{mm^3}{mg}\right]$
K_{d-loq}	Linear sorption coefficient of loq DOC	70.76	$[13.15, 2.32e3]$	$[-]$
$k_{F,hiq}$	hiq DOC growth substrate affinity coefficient of fungi	0.96	$[2e^{-3}, 27.28]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{F,loq}$	loq DOC growth substrate affinity coefficient of fungi	79.17	$[22.54, 147.99]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{F,P}$	Maximum specific rate of pesticide utilization in the absence of growth substrates of fungi	1.02	$[1e-4, 47.26]$	$\left[\frac{1}{d}\right]$
K_{I-B}	Substrate affinity coefficient of insoluble organic matter decomposition kinetics of bacteria and bacterial pesticide degraders	65.19	$[0.16, 2.3e^3]$	$\left[\frac{g}{mg\,C}\right]$
K_{I-F}	Substrate affinity coefficient of insoluble organic matter decomposition kinetics of fungi	54.24	$[0.07, 14.5e3]$	$\left[\frac{g}{mg\,C}\right]$
$k_{m-B,hiq}$	hiq DOC maintenance substrate affinity coefficient of bacteria	652.74	$[0.69, 8.0e^4]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-B,loq}$	loq DOC maintenance substrate affinity coefficient of bacteria	221.47	$[4.87, 1.3e^3]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,hiq}$	hiq DOC maintenance substrate affinity coefficient of bacterial pesticide degraders	269.72	$[0.90, 10.9e^3]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,loq}$	loq DOC maintenance substrate affinity coefficient of bacterial pesticide degraders	1365.72	$[4.55, 3.8e^5]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,P}$	pesticide maintenance substrate affinity coefficient of bacterial pesticide degraders	679.53	$[0.252, 1.1e^4]$	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{r-B,hiq}$	Inhibition coefficient of bacterial activity in response to hiq DOC	0.432	$[0.06, 8.31]$	$\left[\frac{mg\,C}{g}\right]$
$k_{r-B,loq}$	Inhibition coefficient of bacterial activity in response to loq DOC	1.25	$[0.25, 9.74]$	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,hiq}$	Inhibition coefficient of bacterial pesticide degrader activity in response to hiq DOC	0.91	$[0.27, 602.62]$	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,loq}$	Inhibition coefficient of bacterial pesticide degrader activity in response to loq DOC	25.09	$[0.43, 26.57]$	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,P}$	Inhibition coefficient of bacterial pesticide degrader activity in response to pesticide	4.08	$[0.013, 154.35]$	$\left[\frac{mg\,C}{g}\right]$
$k_{r-F,hiq}$	Inhibition coefficient of fungal activity in response to hiq DOC	$9.41e^{-5}$	$[2.05, 100]e^{-5}$	$\left[\frac{mg\,C}{g}\right]$

$k_{r-F,loq}$	Inhibition coefficient of fungal activity in response to loq DOC	122.20	$[2.41, 2.1e^{4}]$	$\left[\frac{mg\,C}{g}\right]$
$K_{s-F,P}$	Substrate affinity coefficient of fungal co-metabolic pesticide transformation kinetic	0.004	$[3, 63]e^{-3}$	$\left[\frac{mg\,C}{g}\right]$
m_{max-B}	Maximal specific maintenance rate of bacteria	3.08	$[0.93, 181.96]$	$\left[\frac{1}{d}\right]$
m_{max-BP}	Maximal specific maintenance rate of bacterial pesticide degraders	3.94	$[0.01, 133.18]$	$\left[\frac{1}{d}\right]$
μ_{max-B}	Maximal specific growth rate of bacteria	68.60	$[3.02, 2.5e^{3}]$	$\left[\frac{1}{d}\right]$
μ_{max-BP}	Maximal specific growth rate of bacterial pesticide degraders	8.41	$[0.47, 28.91]$	$\left[\frac{1}{d}\right]$
μ_{max-F}	Maximal specific growth rate of fungi	9.42	$[3.32, 16.50]$	$\left[\frac{1}{d}\right]$
q_{max-B}	Maximal specific decomposition rate of insoluble organic matter of bacteria and bacterial pesticide degraders	5.84	$[0.034, 47.13]$	$\left[\frac{1}{d}\right]$
q_{max-F}	Maximal specific decomposition rate of insoluble organic matter of fungi	1.19	$[0.2, 5.87]$	$\left[\frac{1}{d}\right]$
$T_{F,P}$	Co-metabolic pesticide transformation capacity of fungi	4968.68	$[56.56, 1.03e^{5}]$	$\left[\frac{mg\,C}{mg\,C}\right]$
T_{y-F}	Growth substrate transformation capacity of fungi	140.88	$[4.19, 252e^{3}]$	$\left[\frac{mg\,C}{mg\,C}\right]$
$Y_{L,hiq}$	Fraction of the decomposed hiq litter transferred to soil	0.36	$[0.113, 0.701]$	$[-]$
$Y_{L,loq}$	Fraction of the decomposed loq litter transferred to soil	0.84	$[0.722, 0.983]$	$[-]$
Y_{r-B}	Efficiency of insoluble organic matter decomposition by bacteria and bacterial pesticide degraders	0.700	$[0.507, 1.0]$	$[-]$
Y_{r-F}	Efficiency of insoluble organic matter decomposition by fungi	0.990	$[0.982, 0.997]$	$[-]$
$Y_{R-F,P}$	Efficiency of co-metabolic pesticide transformation by fungi	0.515	$[0.298, 1.0]$	$[-]$
$Y_{s-B,hiq}$	Substrate uptake efficiency of hiq DOC by bacteria	0.173	$[0.075, 0.306]$	$[-]$
$Y_{s-B,loq}$	Substrate uptake efficiency of loq DOC by bacteria	0.503	$[0.260, 0.854]$	$[-]$
$Y_{s-BP,hiq}$	Substrate uptake efficiency of hiq DOC by bacterial pesticide degraders	0.500	$[0.205, 0.876]$	$[-]$

$Y_{s-BP,loq}$	Substrate uptake efficiency of loq DOC by bacterial pesticide degraders	0.200	[0.033, 0.771]	[−]
$Y_{s-BP,P}$	Substrate uptake efficiency of pesticide by bacterial pesticide degraders	0.900	[0.835, 0.974]	[−]
$Y_{s-F,hiq}$	Substrate uptake efficiency of hiq DOC by fungi	0.113	[0.011, 0.622]	[−]
$Y_{s-F,loq}$	Substrate uptake efficiency of loq DOC by fungi	0.909	[0.824, 0.952]	[−]
r_{B0}	Initial physiological state index of bacteria	0.058	[0.004, 0.139]	[−]
r_{BP0}	Initial physiological state index of bacterial pesticide degraders	0.296	[0.057, 0.336]	[−]
r_{F0}	Initial physiological state index of fungi	0.116	[0.004, 0.348]	[−]

Table A14 Model parameter symbols, descriptions, base values of parameters of the full PECCAD ODE model (M=12, N=59) that define a pesticide stabilization curve.

Symbol	Description	Value	Unit
a_{max-B}	Maximal specific death rate of bacteria	1.43	$\left[\frac{1}{d}\right]$
a_{max-BP}	Maximal specific death rate of bacterial pesticide degraders	1.24	$\left[\frac{1}{d}\right]$
a_{max-F}	Maximal specific death rate of fungi	2.43	$\left[\frac{1}{d}\right]$
$K_{a-B,hiq}$	Inhibition coefficient of bacterial death rate in response to hiq DOC	9.72	$\left[\frac{g}{mg\,C}\right]$
$K_{a-B,loq}$	Inhibition coefficient of bacterial death rate in response to loq DOC	6.07	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,hiq}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to hiq DOC	21.06	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,loq}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to loq DOC	1.63	$\left[\frac{g}{mg\,C}\right]$
$K_{a-BP,P}$	Inhibition coefficient of bacterial pesticide degrader death rate in response to pesticide	1.97	$\left[\frac{g}{mg\,C}\right]$
$K_{a-F,hiq}$	Inhibition coefficient of fungal death rate in response to hiq DOC	75.24	$\left[\frac{g}{mg\,C}\right]$
$K_{a-F,loq}$	Inhibition coefficient of fungal death rate in response to loq DOC	59.99	$\left[\frac{g}{mg\,C}\right]$
$k_{B,hiq}$	hiq DOC growth substrate affinity coefficient of bacteria	347.46	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{B,loq}$	loq DOC growth substrate affinity coefficient of bacteria	4.91	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,hiq}$	hiq DOC growth substrate affinity coefficient of bacterial pesticide degraders	160.86	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,loq}$	loq DOC growth substrate affinity coefficient of bacterial pesticide degraders	195.22	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{BP,P}$	pesticide growth substrate affinity coefficient of bacterial pesticide degraders	297.78	$\left[\frac{g}{mg\,C\,d}\right]$
K_{d-hiq}	Linear sorption coefficient of hiq DOC	0.008	$\left[\frac{mm^3}{mg}\right]$
K_{d-loq}	Linear sorption coefficient of loq DOC	0.003	$[-]$
$k_{F,hiq}$	hiq DOC growth substrate affinity coefficient of fungi	151.04	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{F,loq}$	loq DOC growth substrate affinity coefficient of fungi	69.83	$\left[\frac{g}{mg\,C\,d}\right]$

$k_{F,P}$	Maximum specific rate of pesticide utilization in the absence of growth substrates of fungi	0.19	$\left[\frac{1}{d}\right]$
K_{I-B}	Substrate affinity coefficient of insoluble organic matter decomposition kinetics of bacteria and bacterial pesticide degraders	90.06	$\left[\frac{g}{mg\,C}\right]$
K_{I-F}	Substrate affinity coefficient of insoluble organic matter decomposition kinetics of fungi	1.34	$\left[\frac{g}{mg\,C}\right]$
$k_{m-B,hiq}$	hiq DOC maintenance substrate affinity coefficient of bacteria	716.51	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-B,loq}$	loq DOC maintenance substrate affinity coefficient of bacteria	223.72	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,hiq}$	hiq DOC maintenance substrate affinity coefficient of bacterial pesticide degraders	178.07	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,loq}$	loq DOC maintenance substrate affinity coefficient of bacterial pesticide degraders	700.94	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{m-BP,P}$	pesticide maintenance substrate affinity coefficient of bacterial pesticide degraders	497.26	$\left[\frac{g}{mg\,C\,d}\right]$
$k_{r-B,hiq}$	Inhibition coefficient of bacterial activity in response to hiq DOC	1.62	$\left[\frac{mg\,C}{g}\right]$
$k_{r-B,loq}$	Inhibition coefficient of bacterial activity in response to loq DOC	9.85	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,hiq}$	Inhibition coefficient of bacterial pesticide degrader activity in response to hiq DOC	2.27	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,loq}$	Inhibition coefficient of bacterial pesticide degrader activity in response to loq DOC	1.64	$\left[\frac{mg\,C}{g}\right]$
$k_{r-BP,P}$	Inhibition coefficient of bacterial pesticide degrader activity in response to pesticide	1.19	$\left[\frac{mg\,C}{g}\right]$
$k_{r-F,hiq}$	Inhibition coefficient of fungal activity in response to hiq DOC	0.05	$\left[\frac{mg\,C}{g}\right]$
$k_{r-F,loq}$	Inhibition coefficient of fungal activity in response to loq DOC	6.56	$\left[\frac{mg\,C}{g}\right]$
$K_{s-F,P}$	Substrate affinity coefficient of fungal co-metabolic pesticide transformation kinetic	6.35	$\left[\frac{mg\,C}{g}\right]$
m_{max-B}	Maximal specific maintenance rate of bacteria	0.41	$\left[\frac{1}{d}\right]$
m_{max-BP}	Maximal specific maintenance rate of bacterial pesticide degraders	3.34	$\left[\frac{1}{d}\right]$
μ_{max-B}	Maximal specific growth rate of bacteria	55.89	$\left[\frac{1}{d}\right]$

$\mu_{\max-\mathrm{BP}}$	Maximal specific growth rate of bacterial pesticide degraders	1.20	$\left[\frac{1}{d}\right]$
$\mu_{\max-\mathrm{F}}$	Maximal specific growth rate of fungi	2.46	$\left[\frac{1}{d}\right]$
$\mathrm{q}_{\max-\mathrm{B}}$	Maximal specific decomposition rate of insoluble organic matter of bacteria and bacterial pesticide degraders	3.37	$\left[\frac{1}{d}\right]$
$\mathrm{q}_{\max-\mathrm{F}}$	Maximal specific decomposition rate of insoluble organic matter of fungi	1.45	$\left[\frac{1}{d}\right]$
$\mathrm{T}_{\mathrm{F,P}}$	Co-metabolic pesticide transformation capacity of fungi	4971.72	$\left[\frac{mg\,C}{mg\,C}\right]$
$\mathrm{T}_{\mathrm{y-F}}$	Growth substrate transformation capacity of fungi	0.002	$\left[\frac{mg\,C}{mg\,C}\right]$
$\mathrm{Y}_{\mathrm{r-B}}$	Efficiency of insoluble organic matter decomposition by bacteria and bacterial pesticide degraders	0.62	$[-]$
$\mathrm{Y}_{\mathrm{r-F}}$	Efficiency of insoluble organic matter decomposition by fungi	0.98	$[-]$
$\mathrm{Y}_{\mathrm{R-F,P}}$	Efficiency of co-metabolic pesticide transformation by fungi	0.44	$[-]$
$\mathrm{Y}_{\mathrm{s-B,hiq}}$	Substrate uptake efficiency of hiq DOC by bacteria	0.25	$[-]$
$\mathrm{Y}_{\mathrm{s-B,loq}}$	Substrate uptake efficiency of loq DOC by bacteria	0.96	$[-]$
$\mathrm{Y}_{\mathrm{s-BP,hiq}}$	Substrate uptake efficiency of hiq DOC by bacterial pesticide degraders	0.31	$[-]$
$\mathrm{Y}_{\mathrm{s-BP,loq}}$	Substrate uptake efficiency of loq DOC by bacterial pesticide degraders	0.28	$[-]$
$\mathrm{Y}_{\mathrm{s-BP,P}}$	Substrate uptake efficiency of pesticide by bacterial pesticide degraders	0.45	$[-]$
$\mathrm{Y}_{\mathrm{s-F,hiq}}$	Substrate uptake efficiency of hiq DOC by fungi	0.72	$[-]$
$\mathrm{Y}_{\mathrm{s-F,loq}}$	Substrate uptake efficiency of loq DOC by fungi	0.58	$[-]$
$\mathrm{I}_{\mathrm{hiq}}$	Constant hiq DOC input	0.03	$\left[\frac{mg\,C}{g\,d}\right]$
$\mathrm{I}_{\mathrm{loq}}$	Constant loq DOC input	0.07	$\left[\frac{mg\,C}{g\,d}\right]$
I_{P}	Constant pesticide input	$3e^{-7}$	$\left[\frac{mg\,C}{g\,d}\right]$

UNIVERSITÄT HOHENHEIM

Anlage 2 zur Promotionsordnung der Universität Hohenheim zum Dr. rer. nat.

Eidesstattliche Versicherung gemäß § 7 Absatz 7 der Promotionsordnung der Universität Hohenheim zum Dr. rer. nat.

1. Bei der eingereichten Dissertation zum Thema

   Equifinality, Sloppiness and Emergent Minimal Structures of Biogeochemical Models

   handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Ich habe nicht die Hilfe einer kommerziellen Promotionsvermittlung oder -beratung in Anspruch genommen.

4. Die Bedeutung der eidesstattlichen Versicherung und der strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt.

Die Richtigkeit der vorstehenden Erklärung bestätige ich: Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____ _____

Ort und Datum Unterschrift