Emergent Properties of Heterogeneous Decomposition Networks

by

David C. Forney III

B.S. Mechanical Engineering, University of Delaware (2003)

S.M. Mechanical Engineering, MIT (2007)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2012

Author ....
Department of Mechanical Engineering
May 8, 2012

Certified by..
Daniel H. Rothman
Professor, EAPS
Thesis Supervisor

Certified by..
Anette Hosoi
Professor, ME
Thesis Chair

Accepted by..
David E. Hardt
Chairman, Department Committee on Graduate Students

Emergent Properties of Heterogeneous Decomposition Networks

by

David C. Forney III

Submitted to the Department of Mechanical Engineering
on May 8, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

Organic matter respiration in natural ecosystems is controlled by a network of biologically, physically, and chemically driven processes. Often it is important to estimate total carbon flux from a degradation system or the decay of mass in the system as a function of time. Historically, mass dynamics are predicted by a compartmental model consisting of various degradation processes. This approach provides a complete picture of mass in the model system, but includes extra information unnecessary for modeling overall mass dynamics. Furthermore, these models quickly become highly parameterized and are kept tangible by reducing the number of processes and system states. This thesis suggests a different approach. I allow the degradation system to be inherently complex, but only consider a simple projection of the system necessary to characterize mass dynamics. Because decomposing organic matter is often described as a heterogeneous continuum which transforms and degrades over a wide range of rates, I model degradation as a network consisting of a large, quasi-continuum of states. The retention of carbon in the network is then estimated by using an eigenvalue projection to calculate the impulse response of the degradation system. For a continuous network, the impulse response can be expressed as a Laplace transform of an "exit rate function." I then pose and solve an inverse problem in order to identify the rates of exit of carbon from decomposing plant matter from across North America. Analysis of the calculated exit rate functions and their associated decay data suggest that plant matter decomposition can often be mathematically approximated by a continuum of parallel processes. Within this approximation, the solution of the inverse problem yields the discovery that exit rate functions are on average lognormal. This result suggests that the overall mass dynamics of complex decay networks often collapse to just two parameters: the mean and the variance of the order of magnitude of exit rates from the network. These parameters are then used to assess the effects of climate and litter chemistry on organic carbon turnover and on rate heterogeneity. I also use observed patterns to explain the effect of natural selection in microbial communities on degradation network kinetics.

Thesis Supervisor: Daniel H. Rothman
Title: Professor, EAPS

Thesis Chair: Anette Hosoi
Title: Professor, ME

Acknowledgments

I first thank my parents for their support and for always providing me with the means to achieve my dreams. They maintained a home and environment which gave me full access to knowledge and information about the world. Anything was indeed possible in such an environment and I will carry those feelings with me for life.

I thank my advisor Dan for providing me with the right tools, approach, and way of thinking for tackling such a difficult, complex and unknown scientific problem. Journeys into problems like this remind me of the journeys the great physicists Copernicus, Kepler, Newton, Heisenberg, Einstein, and many others took over the last centuries. Attacking such problems theoretically with limited information is not for the faint of heart. Finding and accepting success in such endeavors takes a well-groomed mind. I also thank Dan for his continued support. Obtaining funding and a project was THE hardest thing I had to overcome during grad school. Not even thesis writing or paper publishing compared to jumping that difficult hurdle. I'm grateful to have had a home in the Rothman lab where I could just think about science and not about my next meal.

I need to thank the members of the LIDET team for providing me with a huge and easy to use data-set that I extensively used in this thesis. It is an extremely important data-set which will provide futher understanding of ecosystems, ecology, and carbon cycling over the next few decades. I am extremely grateful for the effort, organization and thought that they put into this study.

There have been a few heros in my life. However none of them compare to Rebecca J Hemenway. Her ability to stay in touch with whats most important during our many orbits around the sun amidst the numerous hurdles, walls, and bombs that come her way, and her courage to always rise up to meet the next challenge face to face is absolutely inspiring. Her great sense of play as well as her comprehensive approach to and perspective on issues dear to her and many others makes the achievements here appear as mere goosebumps. No matter what happens in the future, your efforts and passions that I have gotten to know over the recent few years will always inspire me and catalyze similar efforts and passions that reside in me as well.

I also need to thank Chris Follett, for great discussions, especially in the months prior to the submission of this thesis. His integrity, intuition, ability, and calm are exactly what is required to excel at a place like MIT. He is also a hero and has been a great scientific peer and mentor for others as well. Truly both a beauty and a geek.

I thank everyone else who has collaborated and inspired me during my time here in the Rothman group: Yonatan, Alex P, Olivier, Hansjoerg, and Hyman, all of you have a great sense of humor as well as an outstanding approach to science, I wish we collaborated more. Danny, Alex L, Mika, Vid, all of you brought amazing backgrounds, tools, abilities and ways of thinking to the lab. I wish I learned more from you all. Alex L, thanks for teaching me the wonders of the command line. Mike A, and Robert, our time was short but you are both extremely able and fantastic. I'm sure you will be successful. Keep your integrity and keep your nose to the grindstone. Just do what is right; you may make a few or many mistakes along the way but thats OK, success will come if you persist.

I thank Prof. David Parks for offering me a position to teach 2.002, Mechanics of materials, during my first semester. Without that gesture, I may not have attended MIT. I also thank him for a research opportunity in his lab, which would have placed me in an exiting but also entirely different place right now.

I thank Martin Polz and Peko for taking time out of their schedules to to discuss basic questions with me and offer their honest insight during this past year. Your success at the institute and positive attitudes are also very inspiring.

Finally, I need to thank the numerous waves of friends I've had here. Friendships in Boston are constantly in flux as that is the nature of this city. Starting from the original snowriders crew, Joe and Nici, and of course Alvin and Svea, to thirsty folks Craig, DQ, et al., to Khanh and Marcio, and finally to my Euro-Bios for making the last few years here so fun and memorable. Boston is a totally different place with you guys. Some of you remain, but most are gone as your journeys take you elsewhere in this ever-changing and exciting world. Alas, it is now my turn to exit. I will always remember and keep you in mind.

Contents

Chapter 1

Introduction

Biological cells and ecosystems are naturally operating machines which continuously transform and transport energy and mass. Despite advances in our conceptual and quantitative physical understanding of molecular biology and physiology, the dynamics of complex biological systems remain an enigma. In many ways, climate research is proceeding along a similar path to biology. Climate is the result of the interaction of various seemingly stochastic physical processes, highly sensitive the the physical conditions of other processes they are coupled to. While we have increasingly better understanding of the first principles which govern the smaller scale processes of these systems, predicting the behavior of these systems on the macroscale is difficult. As a result, biologists and climate scientists statistically process large datasets in order to better understand relationships within their respective systems and overall dynamics. This type of investigation will become more prevalent as we proceed through the digital/data age.

Respiration in natual ecosystems has multi-scale problems similar to those faced in cellular biology and climate studies; the variation in scale is even more pronounced when considering the larger problem of climate, as climate is coupled with natural respiration, which is driven by microbiology. Our primitive and poor understanding of respiration in natural ecosystems [1] is evident when looking at respiration modules in state of the art ecosystem models [2–6]. Whereas the relationship between plant physiology and macroscopic primary production are well understood [4], connecting our understanding of microbial growth, enzyme kinetics, and soil processes [7–12] to overall respiration dynamics is less straightforward. It is clear that soil organic matter degradation has heterogeneous kinetics [13–17], but characterizing that heterogeneity and the developing appropriate constituitive relations between the kinetics and driving biological, chemical, and physical processes is difficult. The appropriate level of model complexity needed to describe a degradation system is unclear [18].

1.1 Overview of Thesis Results

In order to address these fundamental issues regarding respiration in the carbon cycle, we assume that decomposition results from heterogeneous transformations of organic matter amongst a continuum of states. With this approach we then simplify the degradation continuum by expressing it in terms of a single function (chapter 4). We estimate this function by numerically inverting data from a 10 year decay study of plant matter across North America [19–25] (chapter 3). We find that decomposition can be mathematically described by a distribution of faster (often negligible) serial lag-time processes, followed by a heterogeneous decay that proceeds via a lognormal distribution of rates (chapter 4). This result suggests a simple view of decomposition, where the inherent stochasticity associated with decay results in decay dynamics described by just a few parameters (chapter 2).

Our results also provide information regarding the somewhat unclear constituitive relations between climate, composition, and decay kinetics (chapter 2). For example, respiration rates are expected to increase on average by a factor of 2 for a 10 deg C increase in temperature ($Q_{10} = 2$) [26, 27, 21]. There is a debate however whether this relationship holds across the decay continuum from labilie to refractory processes [28–30]. Although limited to a 10 year study, our results show that changes in climate scale all decay rates in the continuum by roughly the same factor ($Q_{10} = 2.0$) [13]. More refractory rates may increase slightly more with temperature due to Arrhenius kinetics [29], but we find this effect is minor (see chapter 6.) Based on this finding, seasonal or climatic effects on decomposition [31] can be modeled by increasing or decreasing all decay rates in the system equally (chapter 6.) We also find that all other climatic parameters such as precipitation, evapotranspiration, etc. also appear to affect the decay continuum in the same uniform manner. Constituitive relationships betweeen composition and decay dynamics however are different. It is well known that microbial nutrients [32, 33] and the presence of refractory compounds [32, 34, 35] can qualitatively speed up or slow down decay. Our approach however allows for a more specific characterization of these effects on the organic matter degradation continuum. Specifically, we've found that microbial nutrients and refractory compounds such as lignin mainly affect only the faster rates in the decay continuum [13]. Slower rates in the distribution are likely controlled by long term transformation processes independent of nutrient availability and change compound lability.

Our results also address fundamental questions about the ecology of decomposers in ecosystems, such as the effect of natural selection on degradation [21] (chapter 5). We find that natural selection appears to select for communities which are better at degrading naturally occuring plant matter than foreign plant matter, and that this effect of natural selection speeds up decomposition by 28%.

The results of the thesis also address certain persistent kinetic trends in decomposition [13] such as the slow-down of decay rates with time [17, 36], sizes of carbon stores and carbon turnover times that are larger than predicted from initial decomposition rates [24], and mean soil carbon ages which are greater than turnover times [37]. The slow-down and long mean ages of carbon relative to initial decay relate to carbon getting stuck in the slow pathways of the decay network. This quantitative advance in soil carbon modeling provides intuitive answers to the above questions and is consistent with the prevalent qualitative view of degradation as a continuum of organic matter states [10,38–41].

Furthermore, the results of this thesis unify many different linear models of organic matter decomposition (chapters 4, 7), such as multi-pool models [34,42,43,24,25], multi-component transformational models [44–48], continuous transformational models [14,15,10] with linear approximation, and continuous parallel decay models [49,41,50]. The impulse response of all of these models can be characterized by a single function, which represents the 'exit rates' of organic matter from all pathways in the network. The regularized inversion method [51] outlined in chapter 3 is a novel way to estimate to estimate the 'exit rate function' and corresponding dynamic parameters of the underlying unknown decomposition system from decay data.

This thesis is comprised of a series of papers [13, 51, 52] which have been either published (Chapters 2 and 3) or are currently in the process of being submitted (Chapter 4.) The remainder this thesis represents unpublished work.

1.2 Overview of Organic Carbon Degradation

Over 10^7 different types of organic substances [53] comprise the roughly 1800 Gt of carbon in soils [54] and biomass and 750 Gt of dissolved and particulate organic carbon in the oceans [55] and marine sediments [54]. A dead eukaroytic cell may introduce thousands of different organic molecules to the non-living natural environment alone [56]. Because there is no large, continuously increasing quantity of any specific organic compound, biological degradation pathways must exist for all types of these naturally occurring compounds. Many of these pathways are encoded for in the approximately 6Mbp [57] metagenomone of diverse soil decomposer communities which may contain 5000-50000 bacterial species per gram soil [58,59] and up to 10^{11} bacterial individuals [60]. Marine sediments may contain a density of 10000 bacterial species per gram [58]. Together, decomposers convert 75 Gt of terrestrial carbon [61] and 50 Gt of oceanic carbon to CO2 every year [54]. During degradation, organic tissues are broken down to particulate or dissolved organic matter [47], which are then processed microbially to carbon dioxide or converted to microbial compounds and by-products [62] that are subsequently transformed and broken down again [10], eventually resulting in a complete conversion of organic carbon to carbon dioxide.

Quantitatively estimating the rates of these processes is difficult for many reasons: the variety of components initially found in tissues vary in lability [63,64,1,34,65]; as decomposition proceeds, organic molecules interact chemically [32, 66, 1] forming humus and other hard-to-degrade compounds [65,67,47]; microbial processes produce compounds which vary in lability [68,10]; and particulate and dissolved carbon bond and sorb to clays and minerals [69–73] forming organo-mineral complexes that also affect decomposability. All of these phenomena either affect the bioavailablil-

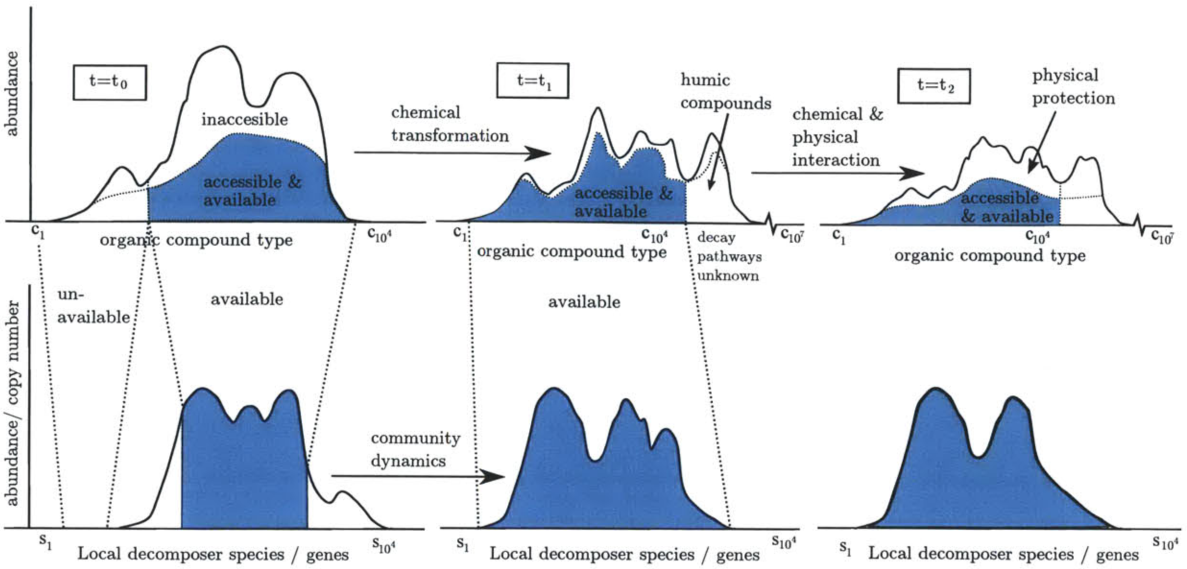

Figure 1-1: Decomposition complexity. Top left: initial distribution of dead plant compounds. Shaded area indicates compounds which can be degraded by initial local decomposer community and are accessible to decomposers. Empty space above the shaded area indicates compounds which are inaccessible to microbes or enzymes because they are embedded in other organic compounds such as lignin [22]. Bottom left: Decomposer community species/genetic diagram. Shaded area indicates initial species with the ability to degrade the substrate. Top center: changes in the substrate composition and reduction in total organic content at time $t1$ due to interaction with other chemical compounds and biological metabolic transformations during the decomposition process. Unavailable area indicates recombination and formation of humic components which may not have a degradation pathway. Bottom center: Changes in local decomposer community to fill the niche of the local substrate. Top right: Further chemical transformational changes in substrate result additional changes in availability. Interaction with surrounding environment, clays, aggregate formation, as decomposition proceeds results in increased inaccessibility. Bottom right: Further changes in decomposer community. Degradation of the blue shaded region proceeds with heterogeneous kinetics.

ity (whether a decomposer with an enzymatic toolkit appropriate for decomposing a given organic polymer is locally available) or the bioaccessibility (whether that compound is physically accessible to a decomposer [74].) The compound may be shielded by other hard to degrade organic polymers which first need to be degraded [75, 65], protected in a tiny pore spaces or in aggregates inaccessible by microbe or perhaps even extracellular enzymes [11], or possibly conformed and attached to surfaces which prevent enzymatic action.

Figure 1-1 depicts a cartoon schematic which qualitatively illustrates the contemporary view of decomposition complexity. The eukaryotic cells of dead plant matter can contain $\sim 10^4$ different organic compounds [56]. Some of these compounds may be unavailable to the initial local decomposer community, or inaccessible to them (as they may be embedded in other organic structures). At a later time t_1, chemical transformations, such as the formation of humic components, and metabolic conversion of compounds can result in a change in available carbon. Changes in the accessible carbon may result from unlocking of shielded compounds and fractionation. Meanwhile the decomposer community may have changed to fill the niche of the substrate. At a later time t_2, physical protection of carbon in small soil aggregates, and the dissolution, diffusion, and interaction of molecules with the local environment may result in further changes in the accessible organic carbon. Furthermore, the accessible and available carbon decays at various timescales, as tissue quality (substrate complexity, substrate nutrient content) [15,10], decomposer community [11,76,77,48],and transport/mobility of nutrients [35,78], enzymes [72,79] , microbes [80,81], substrate, electron acceptor, etc. all affect decomposition. Combined, all of these effects suggest that decay proceeds with stochastic, heterogeneous kinetics [14, 49, 10, 15, 82, 41]. Traditional ways of thinking about these systems often lead to models of increased complexity and high parametrization [83,47,84,85].

1.3 A Common Kinetic Decay Pattern

A simple first-approach for modeling decay with heterogeneous kinetics is to assume it proceeds as a continuous superposition of first-order decays. Using this approach, we have shown that these first-order decay rates are on average lognormally distributed [13] (chapter 2.) This surprisingly simple result suggests that that degradation complexity can be kinetically described by two parameters, a mean and a variance of log rates. Furthermore, this effects of climate (chapters 2 and 6), litter composition (chapter 2, and ecology (chapter 5) on decomposition can be concisely expressed in terms of these two parameters. However, both the origins of the lognormal distribution and the physical interpretation of parallel, first-order decay rates is unclear. To address these questions, I suggest that the continuous superposition of exponential decays derives from the transformation of organic carbon amongst a random network of states (chapter 4). Interaction of the many processes which form this network may result in the observed lognormal distribution(chapter 7. Motivation and background information for the development of this theory is provided in the next section.

1.4 A Different Approach to Decomposition Modeling

As discussed in section 1.2, organic carbon is qualitatively described [38, 10] as a continuum of organic matter states which transform to other states and decay through an assortment of processes occurring on a continuum of timescales. To quantitatively model this heterogeneous system over a wide range of timescales, we make the simplifying approximation that transformations are linear.

The decay system can be modeled deterministically, by a series of linear differential equations [45,18], or stochastically by Markov chains [86] or continuous time Markov processes [87–90]. While I almost exclusively use the deterministic formulation of the system for calculations, the equivalent stochastic framework is useful for visualizing and interpreting the movement and transformation of carbon in a network. Many Markov properties however are not relevant or useful for describing decomposition systems. Typically only average properties related to the mass, residence time, and survival of carbon in the entire degrdation system are of interest, rather than statistical properties of individual transformations in the network. In chapter 4, I present the mathematical approach for describing relevant properties of a system using the deterministic framework. To mathematically introduce a decompostion system as a complex network of transformations, I describe the Markov interpretation of the system next.

A Markov chain consists of a network of many interconnected states. Possible transformations between states are determined by the edges in the network. A example of such a Markov chain is illustrated in Figure 1-2. At each time step, organic carbon in state i may transform to an other state j with the probability p_{ij}. The probability that the carbon remains in state i at the next time step is simply $p_{ii} = 1 - \sum_j p_{ij}$. Markov chains can be expressed in continuous time by expressing the transition probability as a first order transition rate $k_{ij} = P_{ij}/\Delta t$ [87]. Markov chains with continuous time transformations are called continuous time Markov processes. For the case of organic matter degradation, states may be an initial plant compound state, chemically modified state, dissolved state, physically protected state, microbial state, inorganic state, etc. Carbon exits the system when it reaches an exit state, likely an inorganic state resulting from a respiration process.

In Figure 1-2 carbon initially occupying state 1 can be processed only by decomposer 4. Decomposer 4 either respires it to state 7 or converts it to decomposer biomass at state 4. Carbon initially occupying state 2 must first be transformed to state 3, before being processed by decomposers 4 or 5. It may enter the respiration pathway to state 7 or biosynthesis pathways (to states 4 or 5). From the decomposer states, the carbon may pass may transform to soil organic matter state 6 which here may be considered microbial parts or byproducts. From state 6, this organic constituent can be either respired by a microbe or possibly converted back to microbial biomass. The various types of transformations described in Figure 1-2 derive from the cartoon state diagram shown in Figure 1-3.

Organic matter transformations in such a network are related to movement along an energy-decomposition continuum qualitatively depicted in Figure 1-3. States are indicated in italics, and processes in normal typeset. This thermodynamic cartoon qualitatively depicts a few possible paths connecting various types organic matter states in the continuum but includes various major processes associated with decay. An organic carbon atom may enter the system in a complex plant polymer state. Various enzymatic processes may partially or fully degrade that polymer to its monomeric constituents, suitable for microbial respiration. Partially degraded polymers may recombine with other organic or inorganic compounds, thus affecting lability and availability to microbial decomposers. Carbon uptake by decomposers proceeds in two directions. 1) It may be oxidized fully or partially to an inorganic carbon state or partially oxidized microbial byproducts [62]. 2) Or it may proceed along a biosynthesis pathway converting the carbon back into a more reduced complex biopolymer [62]. This simple cartoon only depicts a few states of organic matter.

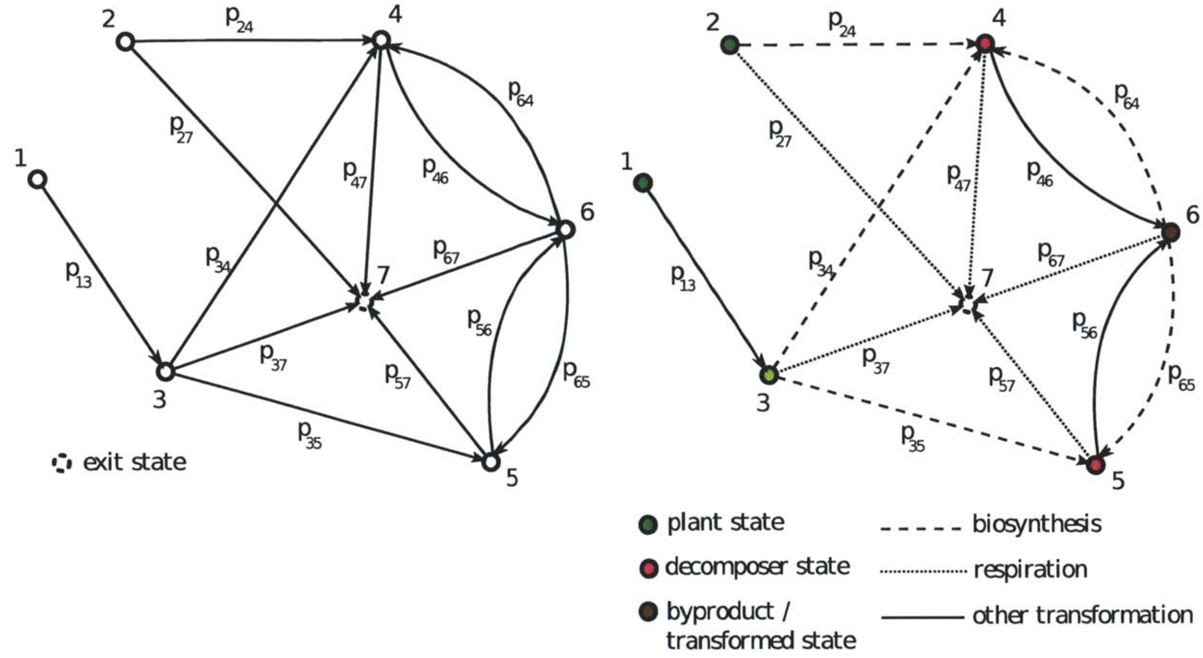

Figure 1-2: Example of a decomposition system expressed in terms of a Markov chain of seven states. Left: Markov chain with state transition probabilities p_{ij} between states i and j. Particles exit the system upon reaching state 7. Right: Interpretation of this Markov chain as a simple decomposition network. States 1 and 2 are initial plant matter states, state 3 is a transformed plant matter state, states 4 and 5 are microbial states, state 6 is a microbial byproduct or transformed dead microbial biomass state, and state 7 is the exit state (CO2). In this simple example, all respiration pathways lead to CO2, while all biosynthesis pathways lead to the transformation of organic carbon to microbial constituents.

These states should be considered as part of a continuum of spread over the entire degradation space. Furthermore, soil organic matter likely contains a wide range of chemical potentials with some containing perhaps even more energy than the original plant compounds.

We combine the concepts from figures 1-1, 1-2, and 1-3 when considering a large decomposition network containing many states. Figure 1-4 depicts such a network. The occupied degradation states of the network reflect the distribution of organic compounds present in the system. Transformation rates out of the states reflect varying degrees of accessibility and availability. For example, carbon in less accessible states may have much slower transfer rates than carbon in more accessible states. The edges in the network on the other hand are more related to the decomposer community. The decomposer community adds various microbial compound states to the network, but more importantly it is responsible for the various enzymatic and biological transformations which comprise many of the edges in such a network. A less diverse decomposer community may only be able to transform some of the organic matter states while a more diverse community with a larger metagenomic toolkit can more fully connect the network. As shown in Figure 1-4, one species may be encoded to perform various transformations in the decay network. Multiple different species

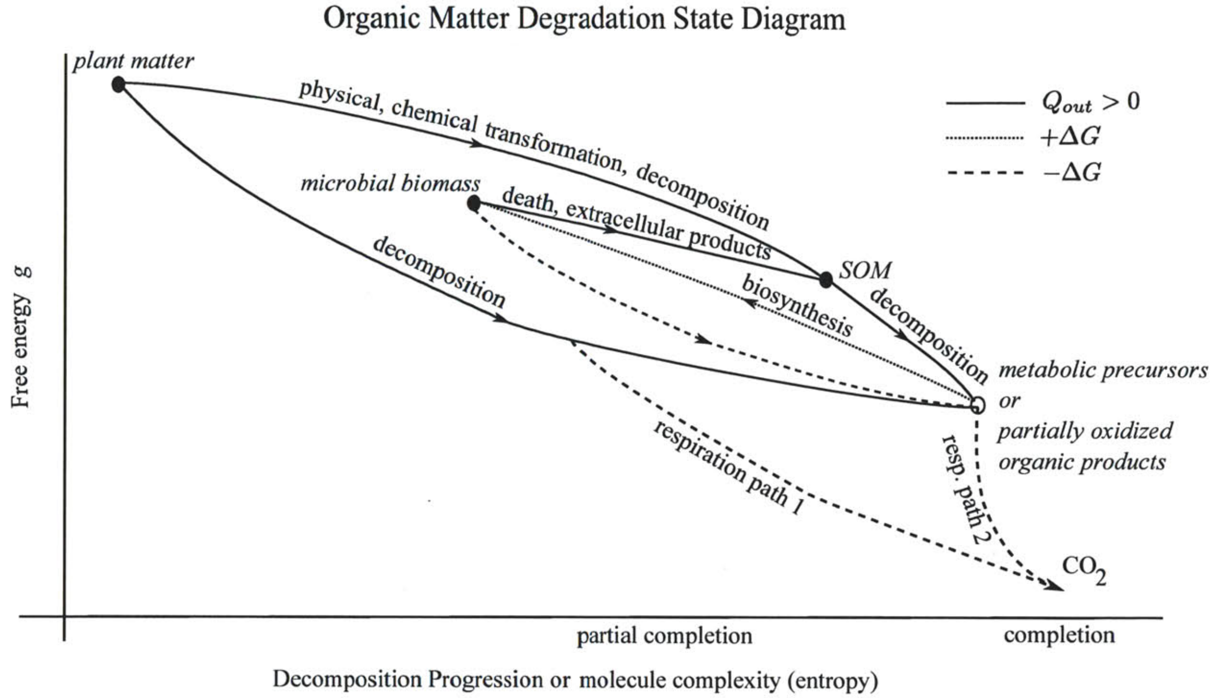

Figure 1-3: Thermodynamic state diagram cartoon. Carbon initially in complex plant polymers can transition to various other partially states through enzymatic and other transformation processes. At various stages of decomposition, degradation pathways may respire the carbon fully to CO_2, partially oxidize the carbon in lack of an electron acceptor and release the carbon to the soil environment, or break down the carbon into precursors for a biosynthetic pathway [62]. Decomposer biosynthesis may result in biocompounds as complex and energy dense as the original plant compounds however are depicted here as slightly less complex for clarity. Some transformations of carbon reduce its lability and it may further transform over time in the soil organic matter pool before being eventually being microbially processed. Many decomposition processes do not take place within the microbial cell. On average those processes are exothermic and that energy is lost as heat to the environment $Q_{out} > 0$. Biosynthesis requires energy $\Delta G > 0$, and cellular respiration releases and stores energy from oxidation of compounds $\Delta G < 0$.

however may be able to transform carbon from state i to j, suggesting that there may be multiple paths (and multiple rates) of transformation between two connected states. We discuss such heterogeneous transfer next.

Consider a specific transformation in a decomposition network from state i to state j. Let state i represent a cellulose polymer which has been partially degraded by an earthworm and now attached in a microaggegate to a clay. State j may represent a hydrolyzed glucose monomer. There are multiple pathways for this transformation to proceed. The glucose may be cleaved by the appropriate hydrolytic enzymes from bacteria X, which is small enough to fit into the microaggregate. It may be hydrolyzed by a biofilm from a fungus Y right next to the aggregate. A randomly passing enzyme diffusing from the background may cleave the glucose. Or perhaps abiotic processes may break the cross-linking hydrogen bonds and, after some long time, some

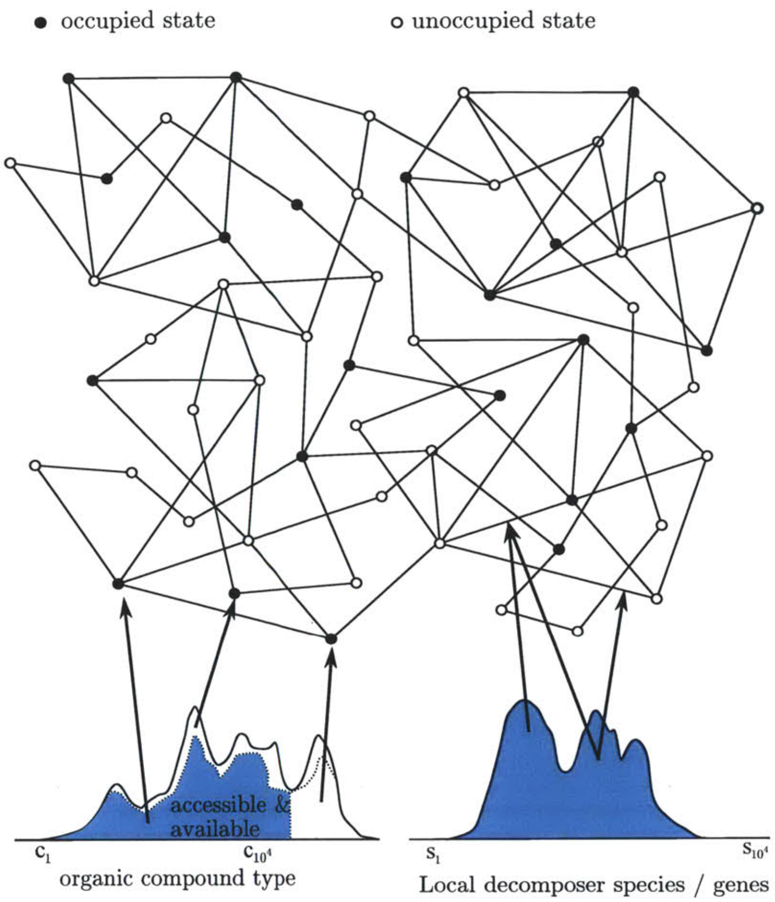

Figure 1-4: A decomposition network.

lucky random fluctuation breaks the covalent bond holding the polymer together. If all processes are first order then the probability of rate of transfer is simply sum of probabilities or rates of all paths between i and j as shown in Figure 1-5A. However, state transformations can proceed in a heterogeneously kinetic manner, and hardly be first order. If the various pathways do not have first order rates and/or interact in a manner that results in a non-exponential distribution of residence times at state i, we can model this behavior as a semi-Markov process with a PH distribution of residence times [88, 90, 91] as shown in Figure 1-5B.

Figure 1-5B shows that transformations with non-homogeneous first order rates and/or an arbitrary, non-degenerate transit time distribution can be equally expressed as a Markov network of sub-states [88]. In other words, any residence time distribution can be represented by an equivalent Markov-process [89, 91]. The residence time distribution associated with a Markov process is called a distribution of phase type or PH distribution [88, 90] and a system comprised of connected states having non-exponential residence time is called a semi-Markov process [89]. Therefore, states with non-exponential residence times can be accounted for by simply increasing the size of the Markov chain.

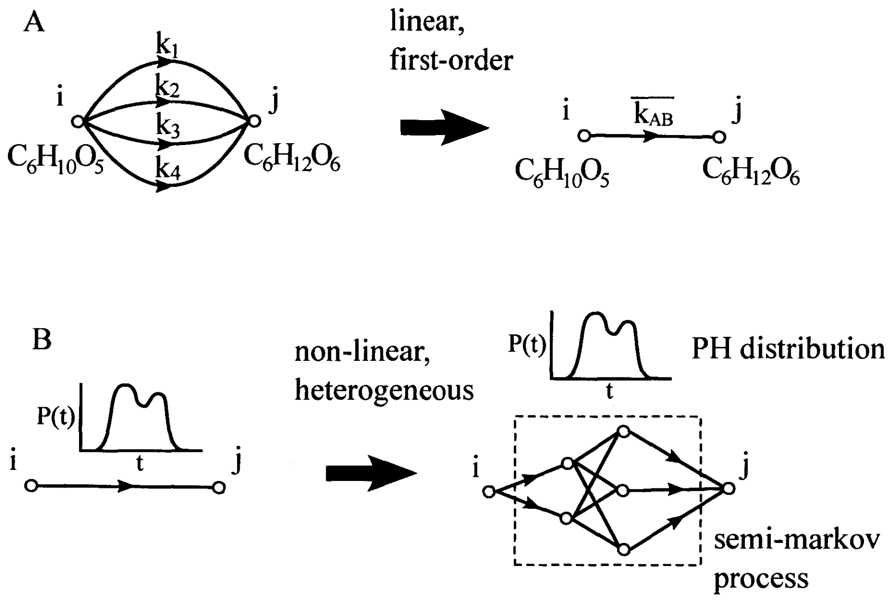

Figure 1-5: State transformations with rate heterogeneity. A: linear, first order rate of transformation. B: Compartment (dashed line) containing heterogeneous rates and having a non-exponential residence time distribution. Any distribution $P(t)$ of residence times t can be approximated by a Markov process. A Markov process consisting of states with non-exponential residence times is called a semi-Markov process.

Similar issues arise when a degradation state is coarsely defined (since the concept of a degradation state itself is vague). If a system is comprised of only a few degradation states, then likely each state is rather coarse, containing heterogeneity. For example, a coarsely defined degradation state may be defined as all soil microaggregates of size $\sim$ 1mm. These microaggregates contain a variety of compounds which may vary in lability suggesting transformations out of this state may proceed with heterogeneous kinetics. This heterogeneity can be easily described with Markov sub-processes containing a number of parallel states. Semi-Markov processes therefore 'take up the slack' in a coarsely defined system.

In this manner, the approach presented in this thesis is general for any decomposition system where state transformations are memory-less. The rate limiting factors, nutrients, enzymes, electron acceptors, physical mixing, etc is accounted for by stochastic and distributed nature of the transformation rates.

Modeling decomposition via compartmental models [45, 92, 93, 83, 47] and small Markov chains [94] is not new. However, prior models only consider networks comprised of just a few states which transform homogeneously at single rates. Because these small networks can easily contain a large number of parameters. The approach presented here, and developed further in chapter 4, quantitatively captures the diversity of timescales and of degradation states which are qualitatively though to be involved in decomposition, but only with the minimum amount of information required to describe mass retention in decomposition systems. This approach and its implementation is discussed in detail in chapter 4 and E. The results of chapters 3 and 2 can also be interpreted in terms of such a stochastic network.

1.5 Applications

Because this is an engineering thesis, and because many people ask me "what are the applications of your work" I've decided to write a little blurb about potential applications, just to provide a framework for those who need it.

Modeling Climate Change

Ecosystem models [2–6] are often coupled with global circulation models (GCM) [95–97] in order to evaluate and estimate climate change. Our primitive understanding of decomposition dynamics as compared to primary production is evident when comparing state-of-the art ecosystem models [2–6]. However, it is difficult to quantitatively model decomposition because we lack fundamental constitutive relations between rates, mechanisms, composition, and environment. While the results of this thesis do not identify those constitutive relations, it provides a plausible framework upon which those relations can be investigated. Incorporation of the models presented here however would require more parametrization before incorporation in coupled GCM-carbon cycle models.

Carbon Economics

Various economic strategies, such as carbon tax, or carbon cap and trade, are currently being evaluated and employed for mitigating the effects of climate change. Soil organic matter may play a role in carbon cap and trade markets, as various forestry and farm management practices [61] (such as till vs. no-till agriculture [48]) may effect the storage of carbon as soil organic matter [12, 11]. Increasing the amount of carbon locked in soils offsets some carbon emissions because that carbon would otherwise have readily been respired to carbon dioxide.

However, before implementing these types of carbon sequestration strategies, metrics regarding their impact and uncertainty [98] need to be established.

The model presented here provides a simple framework for assessing the affects of various practices by evaluating their effect on the exit rate function $v(\ln k)$ discussed in Chapter 4.

Geo-Engineering

Altering the amount of organic carbon stored in farms and forests in a controlled manner is part of a range of geo-engineering topics. Implementation of successful geo-engineering practices will likely proceed through the use of feedback control. The state-space decomposition framework presented here is readily adapted to standard control theory, as shown in Figure 1-6. The equations and variables used to model the system are discussed in detail in chapter 4.

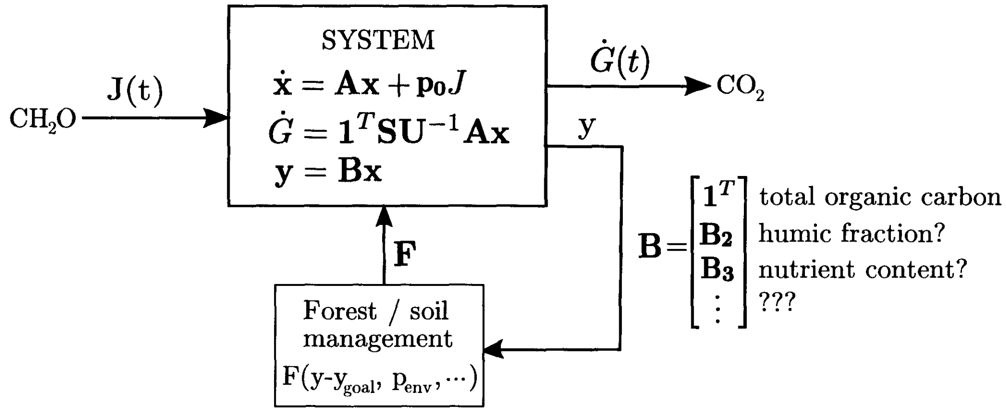

Figure 1-6: Control Diagram. Rate of organic matter input to the system is $J(t)$. $\mathbf{x}$ are the states of the system, $\mathbf{A}$ is the decomposition matrix, G is the total mass of the system, $\dot{G}(t)$ is the mass flux out of the system, $\mathbf{y}$ are observables, and $\mathbf{F}$ are the controlled inputs to the system, which may affect various states $\mathbf{x}$ and network $\mathbf{A}$. Relations for $\dot{x}$ and $\dot{G}$ are derived in Chapter 4.

In order for these practices to be successful, large efforts need to be made on two fronts 1) identifying an appropriate set of important observable soil properties, or metrics $\mathbf{y}$, which will be used as key indicators of the state of the system, used to make management decisions, and to assess whether targets are being met. As suggested in Figure 1-6, a few important observables other than the mass of carbon in the system $\sum_i x_i$ could be soil lignin concentration, N,P,S,K concentrations (based on the results of chapter 2), mass fraction of humic substances, and possibly a number of other observable properties. 2) Identifying successful control strategies $F(y)$ for achieving a target level of sequestration. To do so, we need to better understand the relationship between strategies $\mathbf{F}$ and the degradation system $\mathbf{A}\mathbf{x}$.

1.6 Types of Decay Systems Investigated

Because the wide range of time scales makes it impossible to directly measure decay over all phases of decomposition, here we focus on plant litter decay and early transformations to young soil organic matter. Specifically, we analyze measurements from the Long Term Intersite Decomposition Experiment Team (LIDET) study [21,24,22,25,99]. Litter decay and transformation can be considered as movement along part of the underlying, larger soil organic matter decomposition network. By inverting litter decay data, we are characterizing decay process from a subsection of the soil decay network.

1.6.1 Data

The main respiration study used in this work comes from the Long Term Intersite Decomposition Experiment Team [19] study which took place between 1990 and 2002. During this study over 10000 litter bags of roots, needles, leaves and wood dowels from 27 different vegetation species were planted at 28 locations across North and Latin America. Each data point consisted of a substrate/location pair removed at a given time. Typically there were four replicates of each data point. A data set from the LIDET study consists of the averaged replicates of a unique substrate/location combination. After removing flagged data, only data sets with at least 6 data points (including the time = 0 point) were considered. Furthermore we disregarded a few data sets having negligible mass loss or considerable mass accumulation during the study. With these constraints, we analyzed 237 data sets from 26 different locations and 13 different substrates. The LIDET study also included wet chemical data regarding each substrate as well as environmental properties of each sited.

1.6.2 Role of litter in the carbon cycle

Nearly half of the Earth's primary production of organic carbon occurs terrestrially [73]. About 30% of terrestrial net primary production results in forest litter comprised of leaves, flowers, twigs, fruits, and fine roots which have an average turnover time of 1.5 years. Another 40% terrestrial net primary production occurs in non-forest ecosystems many of which are also provide for biomass with short turnover times. The remaining 30% results in the production of forest wood which stores carbon for an average of 50 years [73]. Although only a small fraction of organic matter will becomes soil organic matter pool, respiration controls the fraction of net primary production (NPP) which enters the gargantuan 3000 GT-C soil organic carbon store. Additionally, since CO2 in the atmosphere would be consumed in $\sim$ 8 years if the input of atmospheric CO2 stopped, quickly releasing carbon back to the atmosphere is a key process that keeps atmospheric CO2 high enough for current levels of photosynthetic carbon fixation [32].

1.7 Outline

Chapter 2: Common Structure in the Heterogeneity of Plant Matter Decay

This chapter applies the method of regularization described in Chapter 3 to data from the LIDET database. We find that plant matter decomposition rates are on average characterized by a log-normal distribution. We interpret the effects of climate and litter chemistry on decomposition by showing how climatic and compositional parameters affect the heterogeneity of decay rates. We also provide an analytical formula for estimating the mean residence time of carbon in dead organic matter.

Chapter 3: Inverse Method for estimating respiration rates from decay time series

This chapter presents the regularization method used in Chapters 3 and 4 to invert decay time series data. It includes an overview of the types of models typically used to describe organic carbon decay data and we show how inverse methods can be applied to estimate the parameters of two different types of popular decomposition models.

Chapter 4: Carbon Transit Through Degradation Networks

This capstone chapter unifies the work from Chapters 2 and 3 as well as other common decomposition models in terms of a continuous Markov chain of degradation states. I assume that organic matter is distributed amongst a continuum of states which transform and decay at a continuum of rates and calculate the retention of carbon in these types of systems. I show how the rate distribution models of chapters 2 and 3 are simply the eigenvalue projections of the initial partitioning of mass in such a network. These projections are called exit rate functions and represent the minimum amount of information required to describe carbon storage in an organic matter decomposition network. I calculate the exit rate functions for 232 LIDET datasets and analyze the results. I show that a lognormal exit rate function depicts decay almost as well as more general exit rate functions.

Chapter 5: Interpreting Ecology from the LIDET study: The homefield advantage

In this chapter I compare the degradation of 10 litters at the location where they were picked (home) to the degradation of those 10 litters abroad. I find that there is a homefield advantage which speeds up degradation by about 30%. I use these results to interpret the effect natural selection on the kinetics of a degradation network.

Chapter 6: The effect of Climate fluctuations on Decay.

This chapter discusses the sensitivity of decay to climatic and environmental fluctuations. I interpret the meaning of rates and the parameter μ when decay fluctuates over time, but measurements are only taken periodically at seasonal intervals. I then show how degradation dynamics change as a function of climate and provide suggestions for implementing the exit rate function approach in large-scale models.

Chapter 7: Conclusion

Here, I discuss how to interpret the meaning of positively constrained exit rate distributions $\rho(\ln k)$ for the case of the network model. I reflect on the major accomplishments of these results. I also discuss how this work combines other models of decomposition into one unifying framework.

Appendices

Appendix A: Mean residence times and ages of particles in linear-time-invariant systems

Here, I derive the residence time distribution, age distribution, mean residence time and mean age for particles which enter any linear reservoir with time-invariant properties. I evaluate those distributions and means when the system can be described by the Laplace transform of an exit rate function. I then provide solutions for those means and distributions when the exit rate function is a lognormal distribution. In order to evaluate the transient dynamics of such a system, for example after a climatic disturbance, I calculate the rise time of a lognormal LTI system after a step disturbance on the parameters μ and σ. The rise time is proportional to the turnover time of the disturbed system.

Appendices B and C: approximations of the Laplace transform of the lognormal

Because the Laplace transform of the lognormal distribution has no closed form solution, this appendix provides and investigates various approximations. Particularly important is the moment closure approximation, commonly suggested to well approximate the lognormal distribution with just a few terms. I find however that the moment expansion requires many terms if $\sigma > 1$.

Appendix D: Apparent decay rate $K(t)$ associated with lognormally distributed rates.

This appendix investigates how the apparent first order reaction rate $K(t)$ changes as a function of time when rates are lognormally distributed. The first order instantaneous rate is defined as $K = -\frac{dg}{dt}/g$ and is important because it can be considered the apparent reaction rate at a time t. This rate characterizes how decay slows down with time. I also discuss how the scale of measurements are important when interpreting dynamical properties like $K(t)$.

Appendix E: Pathway estimation

In this appendix I provide a method for estimating a decomposition pathway based on a decay time series. Similar to chapter 4, I provide an algorithm for calculating the best fitting multi-pool model with negative components. I then provide suggestions for estimating the architecture of the most likely decay pathway based on plausible physical and biological constraints. This method, like the method presented in chapter 3 for estimating the best fitting multi-pool model is highly sensitive to noise. While it is not very appropriate for describing litter decay data, it could be a powerful method for describing less noisy data from other systems.

Appendix F: What is noise?

In this appendix I consider the possibilities that noise in litter decay data comes from random disturbances or that noise reflects the actual heterogeneity in decay rates which vary from bag to bag. I use these results to interpret the kinetic heterogeneity of estimated rate distributions $\rho(\ln k)$. I suggest another way of estimating decay rate distributions based on maximum likelihood calculation of observing the experimental data.

Appendix G: Microbial woodshop

This is simply a fun thermodynamic model of decomposition which places the work of this thesis in perspective.

Chapter 2

Common Structure in the Heterogeneity of Plant-Matter Decay[1]

Abstract

Carbon removed from the atmosphere by photosynthesis is released back by respiration. Whereas some organic carbon is degraded quickly, older carbon persists; consequently carbon stocks are much larger than predicted by initial decomposition rates. This disparity can be traced to a wide range of first-order decay-rate constants, but the rate distributions and the mechanisms that determine them are unknown. Here we pose and solve an inverse problem to find the rate distributions corresponding to the decomposition of plant matter throughout North America. We find that rate distributions are lognormal, with a mean and variance that depend on climate conditions and substrate. Changes in temperature and precipitation scale all rates similarly whereas the initial substrate composition sets the time scale of faster rates. These findings likely result from the interplay of stochastic processes and biochemical kinetics, suggesting that the intrinsic variability of decomposers, substrate, and environment results in a predictable distribution of rates. Within this framework, turnover times increase exponentially with the kinetic heterogeneity of rates, thereby providing a theoretical expression for the persistence of recalcitrant organic carbon in the natural environment.[1]

2.1 Introduction

Greater than 90% of the carbon dioxide input to the atmosphere-ocean system each year derives from the natural decay of organic carbon [100, 21]. Decay is heterogeneous in space and time: organic molecules vary in lability [63, 65]; micro-environmental heterogeneity such as the aggregation of minerals in soil and sediments interfere with degradation [69]; humification and repolymerization [65, 12] result in polymers that are difficult to degrade; and decomposer communities are diverse and varied [77, 65]. Physical and chemical changes in local environment [101] can speed up or prevent decomposition altogether. Spatial heterogeneity of soil nutrient concentrations at

---

[1] This chapter has been published as Forney DC, Rothman DH. Common structure in the heterogeneity of plant-matter decay. Journal of The Royal Society Interface. 2012 Apr. [13]

the meter scale [102] may also influence degradation rate heterogeneity. Combined, these diverse effects yield kinetic heterogeneity: older compounds appear to decay at slower rates than younger compounds [103, 17], and carbon stores and their turnover times are larger than predicted from initial decomposition rates [24].

The sizes of the organic carbon stores and their rates of turnover are required for quantifying feedback between climate and the carbon cycle [95,96] in order to predict changes in carbon dioxide levels and climate [100, 21, 104, 101]. Because decay time scales vary widely, from minutes to millions of years, estimates of carbon stocks and turnover times require knowledge of all decay rates, from fast to slow [45, 18]. However, rate distributions and the mechanisms that determine them remain unknown. Identification of rate distributions should provide insight, not only for predictive purposes [10,39,18] but also for understanding the ecological dynamics [105] of decomposition.

Because the wide range of time scales makes it impossible to directly measure decay over all phases of decomposition, here we focus on plant litter decay and early transformations to young soil organic matter. Specifically, we investigate measurements from the Long Term Intersite Decomposition Experiment Team (LIDET) study [21,24,22,25,99]. This study monitored the decomposition of 27 different types of litter, including needles, leaves, roots, wood, grass, and wheat distributed amongst 28 different locations across North America ranging from Alaskan tundra to Panamanian rainforests. Litter was collected and then re-distributed in litter bags at different sites in order to investigate the effect of composition, ecosystem, and climatic parameters on decomposition. Litter bags were collected and analyzed at least once per year for up to ten years. We show how to estimate kinetic heterogeneity from these observations of decay. We also analyze how these rate distributions are related to climatic conditions and litter composition.

The remainder of this paper is as follows. In section 2, we pose and solve an inverse problem to find the rate distributions corresponding to the decomposition of plant matter from the LIDET study. We then show in Section 3 that the distributions are lognormal on average. Subsequently, in section 4, we show how the two parameters of the lognormal distribution depend on composition and environment. In section 5, we derive the relation of these parameters to the turnover time of carbon stocks. These results show that turnover times grow exponentially as the heterogeneity of rates increase, thereby highlighting the dependence of carbon stocks on their slowest rates of decay [18].

2.2 Disordered Kinetics

Organic matter decomposition may be viewed as the relaxation of a kinetically disordered system. In this section we specify a model for the influence of disorder on decay. We then describe how we invert it to obtain a distribution of decay rates.

2.2.1 Model

We suppose that decay rate constants k derive from stochastic reactions between heterogeneous substrates and ecological communities in a random environment. We describe this scenario using a "static" model of "disordered kinetics" [106–108]. In this model, the mass $g(t)$ is a decreasing function of time that derives from a continuous superposition of exponential decays e^{-kt} weighted by the probability $p(k)\mathrm{d}k$ that the rate constant k is present at the onset of decay. Given these assumptions, decay proceeds as

$$\frac{g(t)}{g(0)} = \int_0^\infty p(k)e^{-kt}\mathrm{d}k \tag{2.1}$$

where $p(k) \geq 0$ and $\int_0^\infty p(k)\mathrm{d}k = 1$. Models similar or identical to equation (2.1) have been previously employed to describe organic matter decay [49, 50, 39, 109, 14, 18, 10, 41, 40, 9]. When the distribution $p(k)$ is discrete, the integral in (2.1) becomes a sum known as a "multi-G" or "multi-pool" model [22,42,34]. Although equation (2.1) lacks detailed mechanisms of the processes involved in decomposition, its simplicity and common application suggests that it is a reasonable first attempt at characterizing decomposition dynamics. Dispersion of the rates k in this model are likely associated with variations in the quality of plant matter compounds [33, 8] which range from highly labile simple sugars to more refractory lignin, waxes and phenolic compounds [34]; local spatial heterogeneity in soil moisture and nutrients [102]; chemical transformation of compounds [12]; and decomposer and metabolic diversity [76, 110]. Rather than attempting a detailed characterization of these individual mechanisms, we simply seek the distribution of rates associated with the minimal description of decomposition given by equation (2.1).

Although equation (2.1) represents a system of parallel, steady decays, decomposition also involves temporal disturbances and serial processes. However, serial transformation processes can be mathematically expressed as parallel decays [18]. Regarding temporal fluctuations, we interpret the steady distribution $p(k)$ as the probability that decay occurs at an effective first order rate k that is averaged over seasonal and other disturbances [18]. We also note that if the difference between the time scales of two serial processes is large, the system effectively relaxes at the time scale of the slower process. For example, the degradation time scale of a particle attached to a mineral surface may be much larger than the duration of the transient period before attachment; similarly, the timescale of humification is likely short relative to the lifetime of the slowly degrading humic substance [18]. Decomposition may be approximated as proceeding initially from the mineral-associated or humic state [18]. A consequence of a parallel decay model is that resulting decays $g(t)$ are convex (concave-up). Specifically, any completely monotone decay $g(t)/g(0)$ can be described by a linear superposition of rates weighted by a probability density function $p(k)$ [111].

We also note that the "random rate model" [106] represented by equation (2.1) has been commonly used to solve problems involving heterogeneous relaxation in other fields. Examples include NMR spin decay [112, 113], protein state relaxation [114], as well as dielectric, luminescent, and mechanical relaxations [106,108,107].

2.2.2 Inverse problem

Under certain physical conditions, distributions $p(k)$ of reaction rates can be calculated analytically [49, 50, 106] and evaluated by comparing $g(t)$ to experimental data. However, given the complex nature of decomposition, purely physical models may not be appropriate. We therefore seek to identify the distribution that best fits the data without resorting to assumptions beyond those implied by equation (2.1). Once the best distribution is found, physical reasoning then allows identification of mechanisms that can generate this distribution.

Mathematically, equation (2.1) is a Laplace transform and $p(k)$ can be found from its inverse. However, the inverse Laplace transform is ill-posed [115], meaning that small changes in the data $g(t)$ can result in large changes in the solution $p(k)$. A standard method to solve such ill-posed problems is to seek solutions $p(k)$ which are minimally "rough" [115]. Here we use Tikhonov regularization [115,116] to identify an optimally smooth $p(k)$ that best fits the data (Section 2.7.2). Such methods have been previously applied to problems of NMR spin relaxation [112] to probe the structure of porous media [117,118] and the properties of biological tissue [113].

2.3 Rates are Distributed Lognormally

We apply this procedure to litter decomposition data from the LIDET study. An example of decay from a LIDET dataset is shown in Figure 2-1A. The corresponding estimate of the rate distribution in logarithmic space, expressed as $\rho(x) = p[k(x)]\mathrm{d}k/\mathrm{d}x$, where $x = \ln k$, is shown in Figure 2-1B. The rate k is rescaled by the period of seasonal forcing (1 year) and is therefore non-dimensional. The good fit of $\rho(\ln k)$ to a Gaussian indicates that the distribution of rates is lognormal, characterized by the parameters μ and σ, where μ is the mean of $\ln k$ and σ^2 is the variance of $\ln k$.

To investigate the extent to which the lognormal distribution applies to the remainder of the LIDET data, we identify the 234 LIDET datasets that contain at least 5 measurements with replicates. These datasets contain 11 different litter types distributed amongst 26 sites. We then employ several tests on each of these 234 datasets to check whether equation (2.1) is an appropriate description of these datasets. First, we find that seven of these datasets show insignificant mass loss between the first and last field measurement, rendering equation (2.1) irrelevant. Six of these datasets are associated with root decay, suggesting that roots can persist for long times in certain conditions.

We next identify the datasets that decay faster than exponentially, counter to the assumption of first-order kinetics and a superposition of exponential decays. We use two tests to identify these datasets. First, we check the curvature of the datasets and find that 9 of the datasets have negative curvature (concave-down). Such superexponential decay cannot be consistent with the Laplace

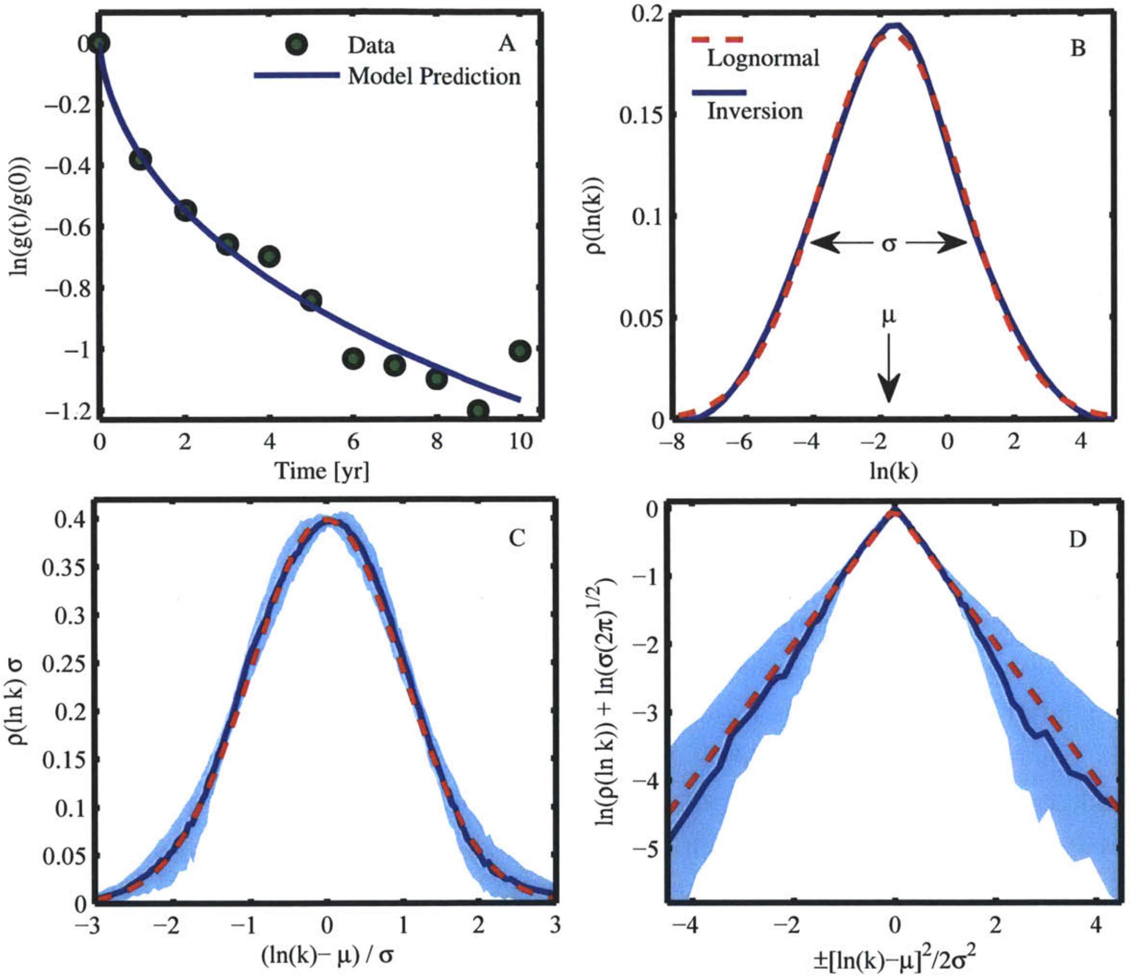

Figure 2-1: Rate distributions of plant-matter decay. **A**: Litter decay from a LIDET data set. Circles are data points. The curve is the predicted decay corresponding to the forward Laplace transform of the solid (blue) curve in B. **B**: Solid curve (blue) is the solution $\rho(\ln k)$ to the regularized inverse problem. Dashed curve (red) is a Gaussian distribution fit to $\rho(\ln k)$. σ^2 is the variance of the Gaussian and μ is its mean. **C**: Whereas Figure B shows just one inversion, the solid curve (blue) is the average of the 182 solutions $\rho(\ln k)$ having non-zero variance, each rescaled by the dataset-dependent parameters μ and σ. Dashed curve (red) is a Gaussian with zero mean and unit variance. The shaded area contains the middle 68% of the numerical inversion results. **D**: Logarithmic transformation of the results of Figure C, where the dashed (red) straight lines indicate an exact lognormal distribution

transform relation (2.1) [111]. These datasets are primarily located at sites having low precipitation, indicating that decay dynamics may be limited by moisture or decomposer activity, rather than substrate availability. Second, we apply our inversion procedure to the remaining datasets and find that three datasets have a significant trend in the residual error. These datasets decay faster than exponentially, but are not concave down. All three of these datasets are associated with wood decay. In summary, our tests disqualify $7+9+3=19$ datasets from further consideration. Further details of the tests are given in Appendix 2.7.1.

Of the remaining 215 datasets, our inversion procedure indicates that 33 datasets are characterized by a single rate constant and decay exponentially. Guided by the result of Figure 2-1B, we then fit a Gaussian to the 182 estimates of $\rho(\ln k)$ exhibiting a non-zero variance and rescale each by the fitted parameters μ and σ. We plot the mean of the rescaled distributions in Figure 2-1C. Although there is scatter and skew among the individual estimates of $\rho(\ln k)$, Fig. 2-1C shows that the mean of the rescaled distributions of $\ln k$ is very similar to a Gaussian distribution. Because the 33 single-rate datasets correspond to a lognormal distribution with zero variance, our results indicate that the lognormal represents the average rate distribution of the 215 datasets for which the model (2.1) applies.

Lognormally distributed variables arise naturally from multiplicative stochastic processes [119]. Here, lognormally distributed rates may result from the multitude of seemingly stochastic requirements for decomposition, such as the presence of water, the presence of an appropriate microbe, the lack of predation, the conditions for expression of hydrolytic enzymes, the encounter of enzymes with the organic matter, etc. [12]. More generally, the probability of completing any task that relies on the successful completion of many subtasks is lognormal [120]. In this context, the lognormal can be viewed as a null hypothesis in which decomposition rates result from the occurrence of a large number of independent decay requirements [12]. Mathematically, if we assume that the probability P of decomposing a parcel of organic matter over a time span Δt is the product of independent probabilities of satisfying various requirements for decay over that interval, then the first-order rate constant $k = P/\Delta t$ becomes asymptotically lognormal as the number of requirements increases. In this manner, the multiplicative stochasticity of a decay system results in the lognormal distribution. This general description suggests that attempts to precisely model the individual mechanisms that stochastically interact to form this broader pattern would be overly complex. It also agrees with the idea that decay rates are the product of many compositional and environmental effects [25].

Previously suggested forms of the rate distribution $p(k)$ are the gamma distribution [49, 18] and the log-uniform distribution [50]. The log-uniform distribution, for which $\rho(\ln k)$ is constant and $p(k) \propto 1/k$ between prescribed limits, approximates the lognormal when $(\ln k - \mu)/2\sigma^2 \ll 1$ [120]. Moreover, its Laplace transform asymptotically approaches the Laplace transform of the lognormal distribution as $\sigma \to 0$. The gamma distribution, however, differs significantly from the lognormal. We find that the lognormal distribution predicts both the data $g(t)$ and describes the inferred rate distribution $p(k)$ better than the gamma distribution for 177 out of 215 datasets (ESM section 2).

We have also compared the lognormal to exponential and multi-pool models. Because our inversion procedure indicates that only 33 of 234 datasets are described by a simple exponential decay, we find that simple exponential decays are generally under-parametrized for describing litter decay datasets, consistent with previous studies [34, 24]. The best fitting type of multiple pool model varies widely amongst the datasets, with no single model type describing all datasets [24, 51]. A universal multiple pool model, containing pools of various types (leached, labile, refractory, inert, etc..), would be over-parametrized. Furthermore, the number of pools and rates associated with each pool are sensitive to noise, as different combinations of pools can represent the same decay [121, 51]. This sensitivity makes understanding the constitutive relationships between pools and environmental and compositional parameters difficult [41].

An advantage of the lognormal is that it parametrizes decay by only two variables, μ and σ. We proceed in the next section to identify relations between the lognormal parameters μ and σ and the climatic and compositional parameters associated with the LIDET study.

2.4 Controls on the Lognormal Parameters

We seek an understanding of the controls on μ (the mean order of magnitude of rates) and σ^2 (the variance of those orders of magnitude). Before analyzing all 215 estimates of these parameters, we identify values of μ and σ which are highly uncertain by disregarding the small fraction of datasets having anomalously long turnover times τ. Assuming a soil carbon store is in steady state with a constant litter input, its turnover time τ is equal to its mean residence time [122], which in the random-rate model (2.1) equals the mean time constant $\langle k^{-1} \rangle$ [41]; thus

$$\tau = \langle k^{-1} \rangle = \int_0^\infty k^{-1} p(k) \mathrm{d}k. \tag{2.2}$$

After evaluating the turnover times associated with all 215 datasets using (2.2), we find that there is a distinct group of datasets associated with excessively long turnover times greater than about > 1000 yr (see Figure 2-5). These datasets contain a significant mass fraction that is effectively inert, having unknown decay dynamics. Extrapolating the kinetics of such slow processes therefore has considerable uncertainty. There are 24 datasets in this outlying cluster. These data are typically associated with root decay at certain locations (see Table 2.3), suggesting that the soils of certain ecosystems can enable the persistence of roots for long times. We do not consider these 24 datasets in our subsequent analysis of μ and σ. Further discussion of these outliers can be found in Appendix 2.7.1.

Due to the nature of the LIDET study, many different litter types were placed at the same site and we therefore have many estimates of μ and σ at each value of temperature, precipitation, and other climatic variables. Similarly, because each litter type was planted at many sites, there are many different estimates of μ and σ for each value of initial lignin concentration, nitrogen concentration, etc. In the following section we study how the average values $\bar{\mu}$ and $\bar{\sigma}$ of the lognormal parameters μ and σ vary with measured independent variables such as temperature, lignin, nitrogen, etc. When analyzing the effects of climatic variables, $\bar{\mu}$ and $\bar{\sigma}$ represent the averages over all litters at each site, and when analyzing the effects of compositional variables, $\bar{\mu}$ and $\bar{\sigma}$ represent the averages over all sites where the litter was deployed. Similarly, $\overline{\sigma^2}$ represents the average variance σ^2 and $\bar{\tau}$ represents the average turnover time, etc. Analagous depictions of the unaveraged data can be found in Appendix 2.7.3.

2.4.1 The mean μ

We first investigate how climate conditions and composition affect $\bar{\mu}$. Figure 2-2A shows a significant positive correlation between $\bar{\mu}$ and temperature. From this trend, we find that the median decomposition rate $e^{\bar{\mu}}$ increases by a factor $Q_{10} = 2.0 \pm 0.3$ (1 standard deviation) with a $10\,^{\circ}\mathrm{C}$ increase in temperature, in agreement with previous estimates [21]. All other measured and synthetic climatic parameters also significantly correlate with $\bar{\mu}$, with the climate decomposition index (CDI) [22, 25] exhibiting the highest correlation (Table 1).

The parameter μ is also related to composition: Figure 2-2B shows that $\bar{\mu}$ decreases as the initial lignin-to-nitrogen ratio (ℓ/N) increases. The observed trend indicates that increases in the lignin concentration, a refractory component of plant matter, are associated with a reduction in $\bar{\mu}$ while increases in organic nitrogen, an important nutrient for microbial decomposers [33, 8], are associated with an increase in $\bar{\mu}$. This is consistent with the use of ℓ/N as a measurement of litter quality [25, 22, 35, 12]. The carbon-to-nitrogen ratio (C/N) and other nutrient measures are also correlated with $\bar{\mu}$ (Table 1). Concentrations of lignin, N, S, P, etc, represent initial values associated with each type of litter. Figure 2B also indicates that needles have lower values of $\bar{\mu}$ than leaves. This effect however may be related to the difference in ℓ/N between the two tissue types.

2.4.2 The variance σ^2

We next investigate the relation of climate conditions to the heterogeneity of decomposition rates, represented by σ. Figure 2-2C shows that temperature has no significant effect on $\bar{\sigma}$. Moreover, $\bar{\sigma}$ is uncorrelated with all climatic parameters monitored in the LIDET study (Table 1); thus climate conditions appear unrelated to σ. We therefore find no evidence from the decadal LIDET data that the Q_{10} of refractory components is significantly different than the Q_{10} of labile components. This supports respiration models such as CENTURY [45, 46], which uses the same temperature and soil moisture factor for each pool of organic matter, independent of lability. We note that if rate dispersion reflects the variation in activation energies of decay processes [29], then Arrhenius kinetics suggest that σ only slightly decreases with temperature over the 35 degree temperature range associated with the LIDET sites. This is consistent with the data presented in Figure 2C, but the wide variation in $\bar{\sigma}$ indicates this trend is not significant and that kinetic heterogeneity is

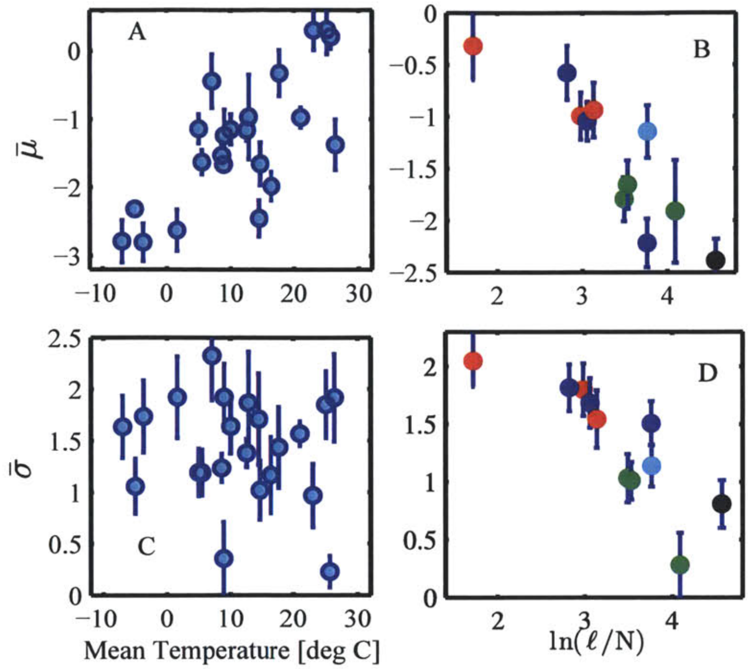

Figure 2-2: Plots of the lognormal parameters $\bar{\mu}$ and σ vs. experimental variables. **A**: $\bar{\mu}$ vs. mean annual temperature. The Spearman rank-correlation coefficient r_s indicates a significant positive trend ($r_s = 0.62$, $P = 0.002$, $N = 22$). **B**: $\bar{\mu}$ vs. the initial litter lignin-to-nitrogen ratio ℓ/N ($r_s = 0.89$, $P = 0.004$, $N = 11$). **C**: $\bar{\sigma}$ vs. mean annual temperature shows no significant relation ($r_s = -0.13$, $P = 0.56$, $N = 22$). **D**: $\bar{\sigma}$ vs. ℓ/N ($r_s = 0.92$, $P < 10^{-5}$, $N = 11$). The color of data points in Figures B and D indicates tissue type: roots (blue), leaves (red), needles (green), wood (black), wheat (cyan). The data in Figures A and C represent 22 sites containing at least 6 different litters each, while the data in Figures B and D represent 11 different litter types planted in at least 4 different locations. Error bars represent one standard deviation of the mean.

controlled by other variables.

Although $\bar{\sigma}$ exhibits no relation to climate, it does vary with composition. Figure 2-2D indicates that $\bar{\sigma}$ decreases as the initial lignin-to-nitrogen ratio ℓ/N increases. Because ℓ/N correlates negatively with $\bar{\mu}$, decreasing the ratio of these components tends to both shift and stretch the rate distribution, increasing the rate constants k of the faster decay processes while the rate constants of slower, more refractory processes are relatively unchanged. Nutrients such as N and S, and to a lesser extent P and K, exhibit similar relations with $\bar{\mu}$ and $\bar{\sigma}$ (Table 1.) Physically these relationships indicate that nutrient limitation is present at early times as faster processes appears to depend strongly on the nutrient content of the litter. Slower, more refractory processes take place at rates likely sustained by the transport and immobilization of nutrients from the surrounding soil [35] and are not nutrient-limited. In fact, increased nitrogen content may inhibit the degradation of transformed plant compounds [12], widening the slow tail of the distribution and increasing σ. Lignin, on the other hand, may reduce the rate constants k of more labile compounds by shielding them via a ligno-cellulose polymer matrix [22], suggesting that ℓ/N measures a resistivity to initial decay.

Table 2.1: Spearman rank correlation coefficients r_s of field experiment parameters vs. $\bar{\mu}$ (left columns) and $\bar{\sigma}$ (right columns). P-values are based on number N of samples used in the rank correlation (final column).

	$\bar{\mu}$		$\bar{\sigma}$		
Parameters	r_s	P	r_s	P	N
precipitation	0.63	2E−3	−7E−3	0.93	22
temperature	0.62	2E−3	−0.13	0.56	22
latitude	−0.51	0.02	0.11	0.62	22
actual evapo-transpiration	0.72	1E−4	0.11	0.62	22
potential evapo-transpiration	0.42	0.05	−0.18	0.41	22
climate decomposition index [22]	0.88	8E−8	−0.02	0.88	22
C/S	−0.71	0.02	−0.87	9E−4	11
C/N	−0.77	8E−3	−0.85	2E−3	11
C/P	−0.45	0.17	−0.48	0.14	11
K	0.65	0.03	0.55	0.09	11
lignin	−0.78	7E − 3	−0.71	0.02	11
lignin/N	−0.89	4E−4	−0.92	0	11
ash	0.68	0.03	0.75	0.01	11
metal	0.63	0.04	0.43	0.18	11
tannin	0.36	0.27	0.25	0.45	11
water soluble	0.52	0.1	0.47	0.15	11
water soluble carbohydrate	0.32	0.34	0.35	0.30	11
cellulose	−0.39	0.24	−0.41	0.22	11
non-polar extractive	−0.23	0.50	−0.37	0.26	11

The effect of ℓ/N on $\bar{\sigma}$ also appears to saturate at low ℓ/N, suggesting that these mechanisms lose control after crossing a threshold [12] of high N or low lignin content is reached. We also observe in Figure 2D that roots and leaves tend to have higher $\bar{\sigma}$ than needles, yet the effect of ℓ/N on $\bar{\sigma}$ appears less strong for roots and wood, both of which decompose underground. Roots and wood do however follow the trend of $\bar{\mu}$ vs. ℓ/N, suggesting that the effect of initial composition may persist over time in roots and wood, effecting a wider portion of their rate distribution, not just the fast rates. This behavior may be a related to components in their tissues, underground decomposition, or both.

Collectively, the results of Figure 2-2 and Table 1 suggest that climate variability changes the median rate of decay, e^{μ}, whereas the variance of decay time scales, σ^2, appears to be a property of the litter sample itself and its relationship to the decomposer community inhabiting it.

2.4.3 Further trends

Table 1 identifies additional correlations between climatic, compositional, and the lognormal parameters. Sulfur, another important microbial nutrient, is highly correlated with both $\bar{\mu}$ and $\bar{\sigma}$. Potassium exhibits a similar trend as well. The causality of the trends in Table 2.1 however is not always clear. For example, ash also has a significant positive correlation with both $\bar{\mu}$ and $\bar{\sigma}$. However, this is most likely explained by the strong rank correlation ($r = 0.89$) between ash and sulfur, as well as a strong correlation between ash and metals which also have a positive correlation with $\bar{\mu}$ and weak correlation with $\bar{\sigma}$. Ash is composed of sulfates, K, P, Ca and other metals [65]. Phosphorus surprisingly does not show as strong a signal as N or S and its large p-values suggests trends with $\bar{\mu}$ and $\bar{\sigma}$ may not be significant. It is possible that the initial phosphorus concentrations may contain error since phosphorous, like N, S, K, and Ca, should have lower concentrations in conifer needles than deciduous leaves [32]; the values of phosphorus measured in the needles and leaves of the LIDET study do not follow this pattern. Metals contain some important rare nutrients for microbial decomposers; we find they are more significantly correlated with $\bar{\mu}$ than $\bar{\sigma}$. The lack of a significant trend for organic compound types (other than lignin) is also surprising, as we would expect water soluble carbohydrates to affect faster decomposition timescales, and cellulose to also play a role in dynamics.

Latitude, used as a proxy for the variability in seasonal temperature, does not show a correlation with $\bar{\sigma}$, indicating that temporal fluctuations in temperature do not contribute to the rate heterogeneity. A comparison of average monthly temperature and precipitation data with $\bar{\sigma}$ also supports this finding. This result provides further evidence that rate heterogeneity is set by non-climatic factors, and that climate scales the timescale of both labile and refractory processes roughly equally.

2.5 Scaling Up to the Carbon Cycle

The heterogeneity of decomposition rates has strong implications for the dynamic properties of carbon stocks. The derivative of equation (2.1) at $t = 0$ reveals that

$$\frac{\dot{g}(0)}{g(0)} = \int_0^\infty k p(k) \mathrm{d}k = \langle k \rangle. \tag{2.3}$$

Equation (2.3) states that the effective initial rate of decay is the mean rate constant [41] because all components are initially present. When $p(k)$ is lognormal,

$$\langle k \rangle = e^{\mu + \sigma^2/2}. \tag{2.4}$$

The mean $\langle k \rangle$ is exponentially greater than the median e^{μ} because of the heavy tail of $p(k)$. A similar amplification acts to exponentially increase the turnover time τ to values much greater than $e^{-\mu}$. Using equation (2.2) and assuming $p(k)$ is lognormal, one finds

$$\begin{aligned} \langle k^{-1} \rangle &= e^{-\mu + \sigma^2/2} \\ &= \langle k \rangle^{-1} e^{\sigma^2}. \end{aligned} \tag{2.5}$$

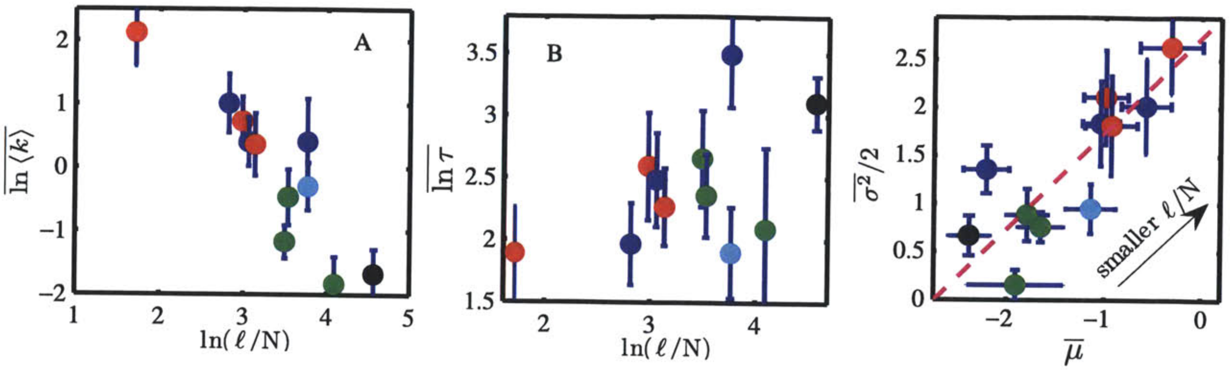

Figure 2-3: Effect of composition on the initial decomposition rate $\langle k\rangle$ and the turnover time τ. The color of data points indicates tissue type: roots (blue), leaves (red), needles (green), wood (black), wheat (cyan). **A**: $\overline{\ln\langle k\rangle}$ vs. the initial lignin-to-nitrogen ratio ℓ/N exhibits a strong negative correlation ($r_s = -0.85$, $P = 0.002$, $N = 11$). **B**: Turnover time $\overline{\ln\tau}$ vs. ℓ/N shows no significant correlation ($r_s = .36$, $P = 0.27$, $N = 11$). **C**: $\bar{\mu}$ and $\overline{\sigma^2}/2$ for each litter type are significantly correlated ($r_s = 0.85$, $P = 0.002$, $N = 11$) The dashed line represents both a constant turnover time $\tau = \exp(-\mu + \sigma^2/2)$ and, by inspection of Figs. 2-2B and 2-2D, the direction of changing ℓ/N. Data points represent 11 different litter types averaged over at least four different locations.

These relations show that rate heterogeneity has a profound effect: $\langle k\rangle^{-1}$ underestimates τ by a factor that grows exponentially with the variance σ^2. As the distribution widens, fast rate-constants weigh heavily on the calculation of $\langle k\rangle$ whereas slower rate-constants set the mean residence time $\langle k^{-1}\rangle$. The upshot is that both the size of organic carbon stocks (proportional to τ in the steady state) and the timescale of the transient response to a disturbance (also related to τ) grow exponentially with the heterogeneity σ^2 of rates. These effects are a consequence of the heavy tail of the lognormal distribution.

We calculate $\langle k\rangle$ and τ for each dataset from our inversion using (2.3) and (2.2) and find the average of the log of their values, $\overline{\ln\langle k\rangle}$ and $\overline{\ln\tau}$ for each litter type. Focusing on the effects of composition, Figure 2-3A shows a strong negative correlation between $\overline{\ln\langle k\rangle}$ and ℓ/N whereas Figure 2-3B shows no significant correlation between the average order of magnitude of turnover time $\overline{\ln\tau}$ and ℓ/N. Physically, these relations reflect the unequal influence of composition on faster and slower rate constants k. Because $\langle k\rangle$ is also the initial decomposition rate, we conclude that the initial ℓ/N exhibits strong control over early decomposition [35, 22]. This influence of initial composition is eventually lost, not only at later times [12] but also in the steady state. Mathematically, these observed trends follow from equations (2.4) and (2.5), given that ℓ/N correlates negatively with both $\bar{\mu}$ and $\bar{\sigma}$ (Figs. 2-2B and 2-2D).

We find that leaves, needles, and roots on average have roughly the same turnover times; 10 yrs, 11 yrs, and 14 yrs respectively. The geometric mean turnover time of all 191 data sets is 11.5 years, but deviations from this characteristic value appear not to be controlled by initial composition. Recall from section 4 that roots may also have uncharacterizably-long residence times in certain locations and these are not analyzed in Figure 3, suggesting a larger departure of root turnover time from needles and leaves. Conditions resulting in extremely persistent root organic matter are unclear (see Appendix 2.7.1). Because the turnover time is unaffected by initial nitrogen concentration, we cannot claim that changes in the nitrogen content of the litter (perhaps through changes in nitrogen deposition) will affect the turnover time of plant matter or carbon storage in soils. It is possible that soil composed of the parent material (as opposed to the LIDET transplant study) may show a different relationship between nitrogen and turnover time. Changes in temperature and precipitation on the other hand affect $\bar{\mu}$ only and therefore do influence turnover time and soil carbon storage. Figure 3A additionally shows a separation in initial decay rate amongst the different litter types, with leaves and roots initially decaying faster than needles. Because the kinetic heterogeneity of roots and needles is wide, one should be especially careful when extrapolating turnover times from short duration decay experiments associated with these tissue types and other litters with high ℓ/N.

Simple patterns emerge from the relationship between composition, μ, σ, $\langle k \rangle$, and τ. The lack of a trend in Fig. 2-3B, combined with equation (2.5), suggests that $-\mu+\sigma^2/2 \simeq$ constant, indicating that μ and σ^2 may be positively correlated under compositional change. Figure 3C shows that σ^2 is indeed correlated with $\bar{\mu}$ across different litter types. Moreover the compositional parameter ℓ/N changes the values of $\bar{\mu}$ and $\overline{\sigma^2}$ roughly along a line of constant turnover time, as expected when $-\mu + \sigma^2/2 \simeq$ constant. Figure 3C also concisely portrays the partitioning of different tissue types in parameter space; needles and wood are characterized by low μ and σ, leaves by high μ and σ, and roots by a range of μ and high σ.

The patterns observed in Figures 3A, 3B, and 3C suggest the following physical interpretation: initial litter composition tends to change the faster rates in the continuum, which affect both μ and σ. The slower rates associated with long-term behavior and turnover time are less related to initial litter chemistry and are more likely determined by soil and microbial community properties. Therefore, during the later stages of litter decay, continued transformation to soil organic matter and its subsequent decay are less a function of the parent material and more a function of semi-transformed compounds and its local interaction with soil [12]. Furthermore, early degradation may be nutrient limited and depend on the nutrient content of the litter, whereas the slower paced degradation of more recalcitrant materials may be sustained by immobilization of nutrients from the surrounding soil. The departure of roots from the trend in Fig 3C, specifically the relative constancy of σ under a change in μ, suggests that the effect of initial composition may persist during root decay or decomposition below ground, influencing the rates of slower processes as well.

2.6 Conclusion

Figure 4 depicts our main findings: *i)* decomposition rates are distributed lognormally; *ii)* environmental change acts as a catalyst which scales all rates similarly, consistent with models such as CENTURY that assign the same temperature and moisture sensitivity across all pools of organic

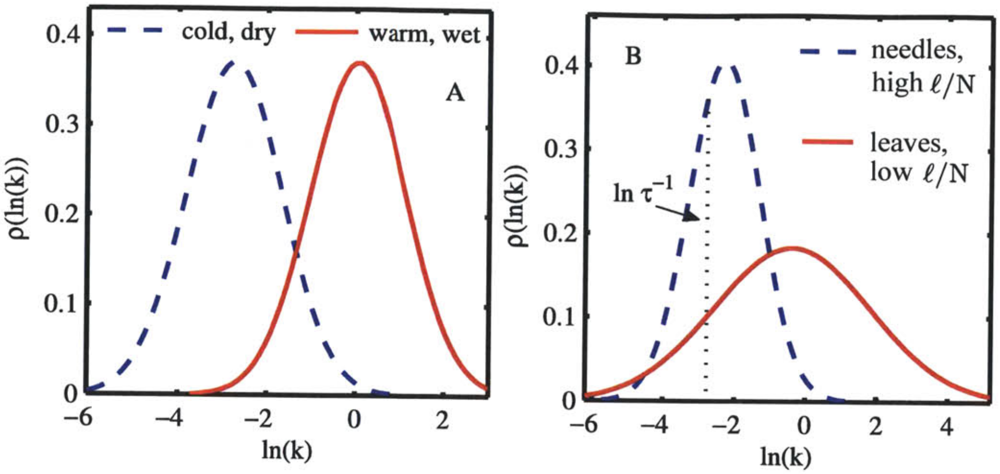

Figure 2-4: Lognormal distributions $\rho(\ln k)$ associated with different climates and plant matter compositions. **A**: Environmental differences tend to shift the distribution along the $\ln k$ axis. Both distributions have a value of σ corresponding to the mean of the data in Fig 2C. The lower value of μ of the (blue) dashed distribution is consistent with values found in colder, dryer climates; the higher value of μ (solid red distribution) is characteristic of warmer, wetter climates. **B**: Faster rates are more sensitive to compositional change, e.g. changing the lignin-to-nitrogen ratio ℓ/N, than slower rates. The dashed blue distribution has values of μ and σ consistent with distributions associated on average with needles or high ℓ/N; the solid red distribution is characteristic on average of leaves or litters with lower ℓ/N. Values of μ and σ are taken from the dashed line in figure 2-3C; thus both distributions result in the same turnover time τ.

matter ; and *iii)* faster processes are more sensitive to litter composition (e.g., ℓ/N, tissue type) than slower processes. The first result, made possible by inverting equation (2.1), identifies the structure of the kinetic heterogeneity associated with decomposition. The second addresses an ongoing debate concerning the temperature sensitivity of decomposition at different time scales [29, 22]. The third result identifies a control for the dispersion of decomposition time scales and shows why composition affects initial decay without changing the turnover time. Each conclusion is separate and independent of the others.

Ecosystem models are often coupled with global circulation models [2–6, 96] in order to provide insight into the climate system. Incorporation of lognormally distributed decay rates in popular ecosystem models and use of the lognormal to precisely predict carbon turnover and storage would require careful parametrizations [45, 46, 12, 96] between the lognormal parameters μ and σ and climatic, soil, and compositional parameters. We have provided a first approach for quantifying these relations. However a more detailed analysis incorporating known mechanisms [12] is required to provide a more comprehensive picture.

We note that the wide range of conditions under which lognormal rates are expected suggests that our results are general, applicable to other degradation processes in natural environments. Evidence of this generality is seen in the decay of marine sedimentary organic matter, which is well described by the quantitatively similar log-uniform distribution [50]. The ubiquity of lognormally distributed degradation rates suggests that a focus on factors that affect rate heterogeneity, rather than specific rates themselves, will lead to a greater understanding — and improved predictions [95, 12] — of the ways in which the carbon cycle interacts with climate.

2.7 Appendix

2.7.1 Data screening

Our analysis uses litter bag data from the Long Term Intersite Decomposition Experiment Team [25, 24, 22, 99]. An important measurement taken during the study is the fraction of original ash-free mass remaining after a given type of plant matter decomposes for a set amount of time in a particular location. Each measurement represents the average of up to four replicates at each site, litter type, and removal time. Litter bags were typically collected and analyzed each year for up to ten years, except for bags at tropical and sub-tropical sites which were more frequently collected at three to six month intervals.

We call a data point the average mass fraction remaining of all replicates of a given plant matter type, site, and duration. If a data point consists of only one litter bag (no replicates), we average that litter bag with the litter bags of the temporally nearest data point (typically one year before or one year after). We call the time series of data points $\tilde{g}(t) = g(t)/g(0)$ associated with a given combination of site and litter type a *data set.* Before analysis, we subject the LIDET data sets to a series of four tests:

1. First we consider only the 234 LIDET datasets which each contained at least five data points (not including time zero) with at least one replicate each.

2. The datasets are subjected to a Mann-Kendall test of trend [123] to check for a significant trend of mass loss between the time of the first data point (not time zero) and the final data point. According to this criterion, seven datasets have no significant decay over this duration and are therefore eliminated from further consideration.

3. The remaining datasets are then checked against the assumption that mass loss is described by the superposition of exponential decays given in equation (1) of the main text and equation (2.15) below. To be consistent with the superposition, any decay $\tilde{g}(t)$ must be completely monotone, meaning that all even derivatives of $\tilde{g}(t)$ must be ≥ 0 and all odd derivatives must be ≤ 0 (see section XIII.4 of Feller [111]). Because, the Mann-Kendall trend test identifies significant decreasing trends, we additionally check only the average sign of the second derivative. To do so, we fit a quadratic function $a + bt + ct^2$ to $\tilde{g}(t)$ and eliminate the nine datasets for which $c < 0$.

4. Of the remaining 218 datasets, three more datasets are eliminated because the inversion algorithm cannot identify a solution $p(k)$ satisfying $p(k) > 0$ and $\int_0^\infty p(k)\mathrm{d}k = 1$ without

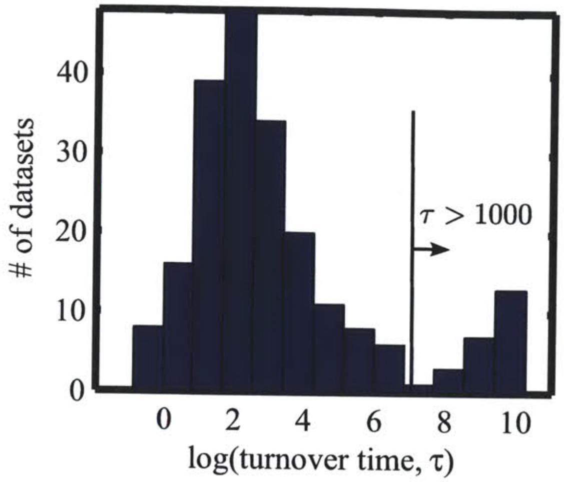

Figure 2-5: Histogram of turnover times of 215 LIDET datasets. The vertical black line has a turnover time of 1000 yr and indicates a clear separation between a main cluster of datasets, and the beginning of a tail which contains extremely long turnover times. We eliminate those datasets to the right of the black line.

having a significant trend in the residual error.

The combined application of these four criteria leaves 215 datasets for analysis. As discussed in the text, we find that the average $\rho(\ln k)$ determined from the inversion of these 215 datasets appears lognormal.

5. Before identifying trends in the parameters μ and σ, we check their values for outliers. Outliers are identified by checking the turnover times of each dataset by substituting the inversion into using equation (2.2). While 155 out of these 215 datasets had turnover times < 50 years, a few datasets were characterized by turnover times extremely large for litter decay, over 1000 years. As shown in Figure 2-5, a histogram of turnover times identified two clusters of datasets; a main cluster of 191 datasets having turnover times less than 1000 yr, and 24 datasets having turnover times greater than 1000 yr.

   For the remaining analysis of the parameters μ and σ, we do not proceed to analyze the outlying cluster of 24 datasets with $\tau > 1000$ for the reasons discussed in the main text.

Trends in data not well described by a superposition of exponential decays

We present the datasets which do not appear well described by equation (2.1) in table 2.3 and find that there are noticeable patterns.

The datasets flagged by test 2 (negligible mass loss) and test 5 (turnover time) are associated with slow degradation and the persistence of organic carbon. Table 2 shows that these datasets are predominantly associated with the decay of *roots* located at the sites CPR, HFR, VCR, NIN,

Table 2.2: Datasets which were flagged by tests 2-5. Datasets numbered 1-19 were not considered in both the inversion and analysis of $\bar{\mu}$ and σ. Datasets 20-44 were included in the inversion but not considered for the analysis of $\bar{\mu}$ and σ. The second column states the code for the location of the dataset, as described in table 3. The third column states the substrate code for the dataset as explained in the table footer. The fourth column states the type of tissue used in each experiment, either needle, leaf, root, wood or wheat. The final column states the reason why each dataset is considered not well described by a superposition of rates, as described in the text of this section.

	site	substrate	tissue	reason
1	CPR	ANGE	root	insignificant mass loss
2	HFR	ANGE	root	insignificant mass loss
3	VCR	ANGE	root	insignificant mass loss
4	NIN	ANGE	root	insignificant mass loss
5	NIN	PIEL	root	insignificant mass loss
6	NWT	PIEL	root	insignificant mass loss
7	SMR	THPL	needle	insignificant mass loss
8	CPR	TRAE	wheat	$g(t)$ is concave down
9	GSF	TRAE	wheat	$g(t)$ is concave down
10	GSF	DRGL	root	$g(t)$ is concave down
11	GSF	PIRE	root	$g(t)$ is concave down
12	JRN	TRAE	wheat	$g(t)$ is concave down
13	JRN	PIRE	needle	$g(t)$ is concave down
14	JRN	THPL	needle	$g(t)$ is concave down
15	SMR	QUPR	leaf	$g(t)$ is concave down
16	LUQ	GOBA	wood	$g(t)$ is concave down
17	BCI	GOBA	wood	trend in residual error
18	BNZ	GOBA	wood	trend in residual error
19	LBS	GOBA	wood	trend in residual error
20	ARC	DRGL	root	$\tau = 2 \times 10^4$
21	ARC	GOBA	wood	$\tau = 3 \times 10^4$
22	ARC	PIEL	root	$\tau = 2 \times 10^4$
23	BNZ	PIEL	root	$\tau = 8 \times 10^3$
24	BSF	ANGE	root	$\tau = 9 \times 10^3$
25	BSF	DRGL	root	$\tau = 1 \times 10^4$
26	BSF	QUPR	leaf	$\tau = 3 \times 10^3$
27	CPR	GOBA	wood	$\tau = 8 \times 10^3$
28	CPR	PIEL	root	$\tau = 2 \times 10^4$
29	GSF	THPL	needle	$\tau = 1 \times 10^4$
30	HFR	PIEL	root	$\tau = 2 \times 10^3$

continued

NWT, ARC, and BSF. While it is known that roots can be highly recalcitrant [12], it is unclear from this data what controls the persistence of roots; there appear to be no common patterns amongst the associated sites; they vary highly in climatic parameters and biome type. Although

Table 2: continued

	site	substrate	tissue	reason
31	JRN	PIEL	root	$\tau = 2 \times 10^4$
32	KBS	PIEL	root	$\tau = 8 \times 10^3$
33	LVW	ACSA	leaf	$\tau = 1 \times 10^4$
34	LVW	PIEL	needle	$\tau = 1 \times 10^4$
35	LVW	QUPR	leaf	$\tau = 1 \times 10^4$
36	NIN	DRGL	root	$\tau = 1 \times 10^4$
37	NWT	ANGE	root	$\tau = 3 \times 10^3$
38	NWT	DRGL	root	$\tau = 1 \times 10^4$
39	NWT	QUPR	leaf	$\tau = 1 \times 10^3$
40	SEV	PIEL	root	$\tau = 2 \times 10^4$
41	SMR	GOBA	wood	$\tau = 1 \times 10^4$
42	SMR	TRAE	wheat	$\tau = 7 \times 10^3$
43	VCR	DRGL	root	$\tau = 1 \times 10^4$
44	VCR	GOBA	wood	$\tau = 3 \times 10^4$
45	VCR	PIEL	root	$\tau = 2 \times 10^4$

The species and common names associated with each code is as follows; ANGE: Andropogon gerardii (Big blue stem), PIEL: Pinus elliottii (Slash pine), THPL: Thuja plicata (Western red cedar), TRAE: Triticum aestivum (Wheat), PIRE: Pinus resinosa (Red pine), QUPR: Quercus prinus (Chestnut oak), GOBA: Gonystylus bananus (Ramin), ACSA: Acer saccharum (Sugar maple), DRGL: Drypetes glauca (Asolillo). τ has units [yr]. Descriptions of the LIDET sites can be found in the ESM Table E1.

all three types of roots (ANGE, DRGL, PIEL) were also planted at the sites: AND, CDR, CWT, HBR, KNZ, LBS, LUQ, LVW, OLY, SMR, and UFL, these sites do not contain a single dataset showing extremely slow root decay. There also appears to be no common patterns amongst these sites.

As discussed in the main text, the datasets filtered by tests 3 (negative curvature) or test 4 (trend in residual error) both identify faster than exponential decays. These were consistently located at sites which had low precipitation or were wood, indicating that decay in dry conditions and decay of wood may not always be substrate-limited.

2.7.2 Regularized inversion

As described in the main text, we assume that a superposition of exponential decays,

$$\tilde{g}(t) = \frac{g(t)}{g(0)} = \int_0^\infty p(k) e^{-kt} \mathrm{d}k \tag{2.6}$$

describes a typical decay experiment, where $\tilde{g}(t)$ is the mass fraction remaining at time t and $p(k)$ is the initial probability distribution of first-order decay rates k. Mathematically, $\tilde{g}(t)$ is the Laplace transform of a given rate distribution $p(k)$. As mentioned in the main text, distributions $p(k)$ have been hypothesized and evaluated by inserting them into (2.15) and comparing the resulting $\tilde{g}(t)$ with data [49,50]. In principle, we could calculate the inverse Laplace transform of $\tilde{g}(t)$ exactly to determine $p(k)$. However, the inverse Laplace transform is ill-posed [124,115], meaning that small changes in $\tilde{g}(t)$ can result in large changes in $p(k)$, effectively resulting in non-unique solutions. We address this problem as follows.

First, we transform the probability density function $p(k)$ to $\rho(x)$, where $x = \ln k$:

$$\rho(x) = p\big(k(x)\big)\frac{\mathrm{d}k}{\mathrm{d}x}, \qquad k = e^x. \tag{2.7}$$

The Laplace transform (2.15) then becomes

$$\tilde{g}(t) = \int_{-\infty}^{\infty} \rho(\ln k)e^{-kt}\mathrm{d}\ln k. \tag{2.8}$$

To solve for $\rho(\ln k)$ from a given set of data $\tilde{g}(t)$, we discretize (2.8) as

$$\mathbf{g} = \mathbf{A}\boldsymbol{\rho} \tag{2.9}$$

where $\mathbf{g}$ is a vector of data points $\tilde{g}(t)$ of length m, $\mathbf{A}$ is the discretized Laplace transform operator of size $m \times n$, and $\boldsymbol{\rho}$ is the solution vector of length n associated with n prescribed intervals of $\ln k$.

Because the system is both constrained and under-determined ($n > m$), we seek the ρ which minimizes the norm of the residual error. In other words, we find the distribution $\boldsymbol{\rho}$ that solves

$$\min_{\boldsymbol{\rho}} \|\mathbf{g} - \mathbf{A}\boldsymbol{\rho}\| \equiv \left(\sum_{i=1}^{n}\left(g_i - \sum_{j=1}^{m} A_{ij}\rho_j\right)^2\right)^{1/2} \tag{2.10}$$

subject to the non-negativity constraint

$$\rho_i > 0, \qquad i = 1 \ldots n \tag{2.11}$$

and the constraint that $\rho(\ln k)$ integrates to unity:

$$\sum_{j=1}^{n} A_{1j}\rho_j = 1. \tag{2.12}$$

This constrained inversion of the Laplace transform, however, remains ill-posed.

Regularization techniques are commonly used to solve under-determined and ill-posed inverse problems [116, 124, 115]. These techniques typically work by attempting to find the "simplest" solution which fits the data signal but not the noise. Here we use a Tikhonov regularization technique [116,115,124], which favors smooth solutions by minimizing both the residual error and a quantitative measure of the complexity of $\rho(\ln k)$. Solution complexity is commonly associated with the roughness or the intensity of fluctuations present in the solution. Here we measure roughness by the squared norm of the first derivative of the solution vector, i.e.,

$$\left\|\frac{\mathrm{d}\rho(\ln k)}{\mathrm{d}\ln k}\right\| = \left(\sum_i \left(\frac{\rho_{i+1} - \rho_i}{\ln k_{i+1} - \ln k_i}\right)^2\right)^{1/2}, \tag{2.13}$$

The problem then becomes

$$\min_{\boldsymbol{\rho}}\{||\mathbf{g} - \mathbf{A}\boldsymbol{\rho}||^2 + \lambda||\mathbf{R}\boldsymbol{\rho}||^2\}, \tag{2.14}$$

where $\mathbf{R}$ is the bi-diagonal first derivative operator (i.e. the discretization of (3.8)) and λ is the regularization parameter, which controls the relative weights of the solution roughness and the residual error norm.

The method proceeds by identifying the value of λ that sets an optimal balance between the residual error $||\mathbf{g} - \mathbf{A}\boldsymbol{\rho}||$ and the roughness $||\mathbf{R}\boldsymbol{\rho}||$. A common approach is to use an L-curve technique [124, 115]. An L-curve is generated by parametrically varying λ and solving (2.14) for $\rho(\ln k)$, obtaining values for the residual error norm $||\mathbf{g} - \mathbf{A}\boldsymbol{\rho}||$ and the roughness norm $||\mathbf{R}\boldsymbol{\rho}||$ for each value of λ. The parametric relationship of $||\mathbf{R}\boldsymbol{\rho}||$ vs. $||\mathbf{g} - \mathbf{A}\boldsymbol{\rho}||$ typically has the shape of an "L". Although each point on the L-curve may be considered an optimal solution for a certain value of λ, the corner of this curve is associated with the regularized solution to the inverse problem. The corner represents the value of λ for which increasing λ strongly increases solution roughness while providing little reduction in residual norm, and also for which decreasing λ greatly sacrifices residual norm with little reduction in roughness.

Before choosing the corner, we check the solution with $\lambda = 0$ in case it is just a single delta function, indicating that only one rate is present during decay. If $>$ .99 of the mass fraction is contained within one rate, we conclude that the simplest solution is an exponential decay having a single rate constant. This occurs for 33 of the datasets.

2.7.3 Unaveraged parameter analysis

We present here the unaveraged data shown in Figures 2 and 3 of the main text. Figure 2-6 shows how the values of μ and σ for the 191 LIDET datasets vary with temperature and the lignin to nitrogen ratio ℓ/N. Figures 2-7A and 2-7B show the unaveraged variation of $\langle k\rangle$ and τ with ℓ/N. Figure 2-7C shows the unaveraged plot of $\sigma^2/2$ vs. μ. While there is much scatter amongst the data, the general trends remain the same as the in Figures 2 and 3 of the main text.

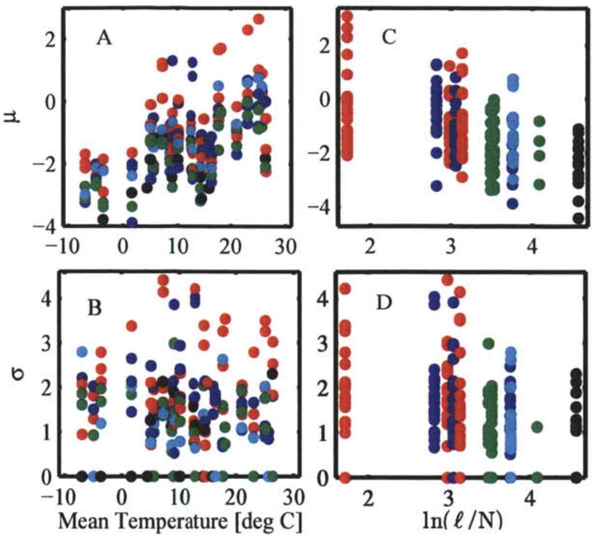

Figure 2-6: Plots of the unaveraged lognormal parameters μ and σ vs. experimental variables temperature and ℓ/N. All 191 datasets are shown in each figure. Colors indicate tissue types. roots (blue), leaves (red), needles (green), wood (black), wheat (cyan). Figure A: μ vs. mean annual temperature. Figure B: μ vs. ℓ/N. Figure C: σ vs. mean annual temperature. Figure D: σ vs. ℓ/N. Comparison with Figure 2 of the main text shows similar trends.

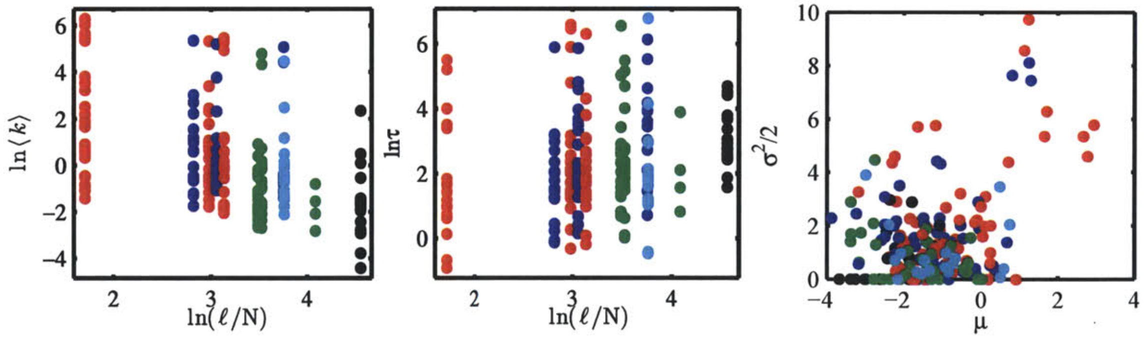

Figure 2-7: Effect of composition on the unaveraged mean decomposition rate $\langle k\rangle$ and turnover time τ. All 191 datasets are shown in each figure. Colors indicate tissue types. roots (blue), leaves (red), needles (green), wood (black), wheat (cyan). Figure A: $\langle k\rangle$ vs. ℓ/N. Figure B: τ vs. ℓ/N. Figure C: $\sigma^2/2$ vs. μ. Comparison with Figure 3 of the main text shows similar trends.

2.7.4 Comparison of the lognormal and gamma distributions

As described in the main text, we assume that a superposition of exponential decays

$$\hat{g}(t) = \int_0^\infty p(k)e^{-kt}\mathrm{d}k \tag{2.15}$$

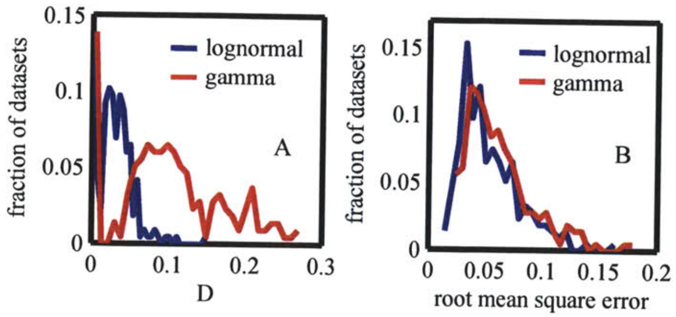

Figure 2-8: Comparison of the Lognormal distribution to the Gamma distribution. A: Kolmogorov-Smirnov D-statistics associated with comparing the inversion to both the normal distribution and $\ln k$-transformed gamma distribution which best fits $\rho(\ln k)$. 215 datasets. B: Mean Squared Error of the predictions from best fitting lognormal and gamma distributions using equation (2.18).

describes a typical decay experiment. Here, we compare a lognormal distribution of rates k,

$$\Lambda(k) = \frac{1}{\sqrt{2\pi}\sigma k} \exp\left(-\frac{(\ln k - \mu)^2}{2\sigma^2}\right), \tag{2.16}$$

parametrized by μ and σ, to a gamma distribution of rates,

$$\gamma(k) = k^{a-1}\frac{\exp(-k/b)}{\Gamma(a)b^a}, \tag{2.17}$$

parametrized by a and b.

We use two statistical methods, a Kolmogorov-Smirnov D statistic and a nonlinear least squares parameter estimation to identify which distribution is a more appropriate description of the underlying litter decay rate distribution.

Kolmogorov-Smirnov D-Statistic

We compare the 215 inversion results to the best fitting gamma and lognormal distributions using a Kolmogorov-Smirnov D-statistic [116]. The D-statistic is a non-parametric measure used to compare two probability distributions [116]. A lower D-statistic indicates a higher likelihood that the data distribution is described by the hypothesized distribution. It is calculated by taking the maximum difference between the cumulative density function, cdf, of an observed distribution (in this case, the inversion), and a hypothesized distribution (the gamma or lognormal). Figure 2-8A shows the distribution of D-statistics which measure the distance between the best fitting Gaussian distribution and the 215 inversion estimates $\rho(\ln k)$. We also identify the $\ln k$-transformed Gamma distribution which best fits $\rho(\ln k)$ and plot its distribution of D-statistics for the 215 inversions. It is clearly seen that the lognormal consistently fits the inversions better than the gamma distribution. The gamma distribution had a lower D-statistic for only five out of 215 datasets.

Error Estimation of Nonlinear Least-Squares Fit

For this comparison, we insert the lognormal (2.16) and gamma (2.17) distributions into (2.15) and estimate their parameters by a nonlinear least squares estimation procedure. We then compare the squared norm of the residual error of the decay predicted by the two distributions by computing

$$\sum_{i=1}^{n}\left(g_i - \int_0^\infty p(k; x_1, x_2)\exp(-kt_i)\mathrm{d}k\right)^2, \tag{2.18}$$

where x_1, x_2 are the two parameters of the lognormal or gamma distributions. We then use the best fitting parameters to predict the decay $g(t)$. The root mean squared error between the lognormal prediction $\mathrm{RMSE} = \sqrt{\sum(\mathrm{g} - \hat{\mathrm{g}})^2/\mathrm{n}}$ is calculated for the 215 datasets, where n is the number of data points in the dataset, $\hat{g}$ is the data, and g is the prediction. The distribution of RMSE values are presented Figure 2-8B. We also calculate the root mean squared error associated with the gamma distribution, and also present its distribution of values in Figure B. The RMSE of the gamma distribution is shifted slightly to the right of the lognormal. We find that for 177 out of 215 datasets, the lognormal has both a lower D-statistic and a lower residual error norm. The lognormal consistently predicts the data consistently better than the gamma distribution, but only marginally given the large noise in the data $\hat{g}$.

2.7.5 LIDET site descriptions

Table 2.3: Descriptions of LIDET sites [25]. MAP: mean annual precipitation, MAT: mean annual temperature: AET: actual evapotranspiration, PET: potential evapotranspiration

Code	Field Site Name	Biome Type	MAP (cm)	MAT (deg C)	AET (cm)	PET (cm)
AND	H.J. Andrews Exper. Forest, OR	Temperate Coniferous Forest	230.9	8.6	76.4	98.2
ARC	Toolik Lake, AK	Arctic Tundra	32.7	-7.0	28.4	42.3
BCI	Barro Colorado Island, Panama	Humid Tropical Seasonal Forest	269.2	25.6	136.8	151.7
BNZ	Bonanza Creek Forest, AK	Boreal Forest	40.3	-5.0	36.0	57.6
CDR	Cedar Creek Natural History Area, MN	Temperate Savanna	82.3	5.5	73.3	102.6
CPR	Central Plains Exper. Range, CO	Temperate Shortgrass Steppe	44.0	8.9	43.0	120.2
CWT	Coweeta Hydrological Laboratory, NC	Temperate Deciduous Forest	190.6	12.5	117.3	135.3
GSF	Guanica State Forest, Puerto Rico	Dry Tropical Forest	50.8	26.3	50.2	142.2
HBR	Hubbard Brook Exper. Forest, NH	Temperate Deciduous Forest	139.6	5.0	71.2	81.7
HFR	Harvard Forest, MA	Temperate Deciduous Forest	115.2	7.1	85.1	104.1
JRN	Jornada Exper. Range, NM	Warm Semi-desert	29.8	14.6	29.2	166.6
JUN	Juneau, AK	Temperate Coniferous	287.8	4.4	49.5	54.4
KBS	Kellogg Biological Station, MI	Temperate Agro Ecosystem	81.1	9.0	70.6	100.7
KNZ	Konza Prairie Research Natural Area, KA	Temperate Tallgrass Prairie	79.1	12.8	74.7	125.0
LBS	La Selva Biological Station, Costa Rica	Humid Tropical Forest	409.9	25.0	169.9	177.3
LUQ	Luquillo Experimental Forest, Puerto Rico	Humid Tropical Forest	336.3	23.0	123.4	125.9
LVW	Loch Vale Watershed, CO	Boreal Forest	109.6	1.6	85.1	108.3
MTV	Monte Verde, Costa Rica	Tropical Elfin Cloud Forest	268.5	17.7	108.4	116.6
NIN	North Inlet (Hobcaw Barony), SC	Wetland	149.1	18.1	120.6	145.6
NWT	Niwot Ridge, CO	Alpine Tundra	124.9	-3.7	64.7	75.6
OLY	Olympic National Park, WA	Temperate Coniferous Forest	153.1	10.0	79.4	104.4
SEV	Sevilleta National Wildlife Reserve, NM	Warm Semi-desert	25.4	16.0	25.2	160.2
SMR	Santa Margarita Ecological Reserve, CA	Annual Grassland	24.0	16.4	23.6	186.0
UFL	University of Florida, FL	Temperate Coniferous Forest	123.8	21.0	116.6	162.1
VCR	Virginia Coast Reserve, VA	Temperate Wetland	113.8	15.0	99.3	121.5

Chapter 3

Inverse Method for Estimating Respiration Rates from Decay Time Series[1]

Abstract

Long-term organic matter decomposition experiments typically measure the mass lost from decaying organic matter as a function of time. These experiments can provide information about the dynamics of carbon dioxide input to the atmosphere and controls on natural respiration processes. Decay slows down with time, suggesting that organic matter is composed of components (pools) with varied lability. Yet it is unclear how the appropriate rates, sizes, and number of pools vary with organic matter type, climate, and ecosystem. To better understand these relations, it is necessary to properly extract the decay rates from decomposition data. Here we present a regularized inverse method to identify an optimally-fitting distribution of decay rates associated with a decay time series. We motivate our study by first evaluating a standard, direct inversion of the data. The direct inversion identifies a discrete distribution of decay rates, where mass is concentrated in just a small number of discrete pools. It is consistent with identifying the best fitting "multi-pool" model, without prior assumption of the number of pools. However we find these multi-pool solutions are not robust to noise and are over-parametrized. We therefore introduce a method of regularized inversion, which identifies the solution which best fits the data but not the noise. This method shows that the data are described by a continuous distribution of rates which we find is well approximated by a lognormal distribution, and consistent with the idea that decomposition results from a continuum of processes at different rates. The ubiquity of the lognormal distribution suggest that decay may be simply described by just two parameters; a mean and a variance of log rates. We conclude by describing a procedure that estimates these two lognormal parameters from decay data. Matlab codes for all numerical methods and procedures are provided. [1]

---

[1] This chapter has been published as Forney DC, Rothman DH. Inverse method for estimating respiration rates from decay time series. Biogeosciences Discussions. 2012 Mar;9(3):3795-3828. [51]

3.1 Introduction

Over 10^7 different types of organic substances [53] comprise the roughly 1800 Gt of carbon in soils [54] and biomass and 750 Gt of dissolved and particulate organic carbon in the oceans [55] and marine sediments [54]. Diverse decomposer communities (up to 6000 bacterial species per gram soil and up to 11000 bacterial species per gram marine sediments [58]) respire these compounds, converting 75 Gt of terrestrial carbon [61] and 50 Gt of oceanic carbon to CO2 every year [54]. During degradation, organic tissues are broken down to particulate or dissolved organic matter [47], which are then processed microbially to carbon dioxide or converted to microbial compounds and by-products [62] that are subsequently transformed and broken down again [10], eventually resulting in a complete conversion of organic carbon to carbon dioxide. Quantitatively estimating the rates of these processes is difficult for many reasons: the variety of components initially found in tissues vary in lability [63,64,1,34,65]; as decomposition proceeds, organic molecules interact chemically [32,66,1] forming humus and other hard-to-degrade compounds [65, 67, 47]; microbial processes produce compounds which vary in lability [68, 10]; and particulate and dissolved carbon bond and sorb to clays and minerals [69–73] forming organo-mineral complexes that also affect decomposability.

This wide range of decomposition processes and mechanisms lead to heterogeneous kinetics, with rates of decay ranging from weeks to thousands of years [17, 36, 37] or greater [103]. Knowledge of rate heterogeneity is important because it tells us how long carbon resides in organic matter. Rates of decay can be related to other dynamic properties such as the turnover times and ages of soil carbon [93, 13, 122, 125, 39, 41]. However, it is difficult to quantitatively model decomposition because we lack fundamental constitutive relations between rates, mechanisms, composition, and environment. The primitive understanding of decomposition dynamics as compared to primary production is evident when comparing state-of-the art ecosystem models [2–6]. Empirically identifying constitutive relations requires 1) mathematical relations underlying the dynamics of decay and 2) estimating dynamic parameters of the model from decomposition data. This paper focuses on the latter problem. Because decomposition dynamics involves multiple time-scales and is highly complex, heuristic models are used. Because degradation can take place over such a wide range of timescales, decay experiments sample only a portion of the decomposition history. As a compromise, we investigate plant litter decay phases and early transformations to young soil organic matter by analyzing a long-term litter decay study that spans up to 10 years.

Current degradation models differ in the way they account for kinetic heterogeneity. They vary in the number of rate pools, mass flow partitioning, and complexity [9]. Models of organic matter decomposition can be classified in terms of three types. The first and most commonly applied model is the multi-pool model [22, 24, 25], otherwise known as multi-compartment [43, 73], multi-component [83, 68], or multi-G [42] models. In these models organic matter is partitioned into one [126] or more [34, 42, 43, 24, 25] pools decaying exponentially at different rates. Pools are suggested to be associated with different compounds present in plant matter [34]. However, the number of rates present, amount of material at each rate, and the value of each rate may vary across different types of organic matter, ecosystems, and climates, and the relations between these parameters are not well understood [22, 41]. The second class of models are continuous parallel models, also called reactive continuum, or disordered kinetic models. In these models, organic matter is described as a continuum of qualities, and decomposition proceeds by a continuous distribution of exponential decay rates [49, 41]. These have only more recently been applied [39, 50, 127, 13]. The third and most detailed type of models are transformational models, which incorporate transformations to other types of soil organic matter, decomposer biomass, humus, etc. These models can be discrete, consisting of a network of pools exchanging carbon with one another [44–48], or continuous [14, 15, 10]. Some transformational models also couple the rates of organic carbon decomposition with nutrient dynamics [7, 10] and may be nonlinear [7, 9]. Transformations in organic matter can also be described by a first-order decay constant which decreases according to a function of time [17, 36, 109].

The three types of models are similar. The multi-pool model can be written in terms of the disordered kinetic model when the continuous rate distribution is comprised of delta functions [13]. The disordered kinetic model can be represented by a transformational model when quality is preserved during transformation [109]. When linear, transformational network models can sometimes be mathematically represented by a multi-pool model [18]. This simplification also occurs when the timescale of transformation is short with respect to the decay timescale of slowly degrading products [13].

Using the disordered kinetic approach, we have recently found that plant matter decay is well described by a lognormal distribution of rates [13] and that the two parameters of the lognormal distribution are qualitatively related to climatic and compositional variables. A distribution of rates provides a unique way to visualize and pinpoint key dynamic information regarding carbon in the system and reveal the timescales of the retention of organic carbon in decomposition systems.

In this paper, we elaborate on the formation and implementation of the regularized inversion method that we employed to identify the best fitting rate distribution [13]. To do so, we first consider a simpler direct inversion of the data which forms the basis for the regularization method. The direct inversion tends to find discrete rate distributions with mass contained at only a few rates. In this sense, the direct inversion is a new procedure for finding the best fitting multi-pool solution without needing to assume how many pools are present. We use the resulting multi-pool solutions to motivate the use of a regularized inverse. We then describe the regularization method in detail and provide an example of its implementation. Because the regularization method identifies a common lognormal pattern [13], we also present a simple procedure for estimating the lognormal parameters from the decay data. This procedure provides a more precise estimate of the best fitting lognormal, resulting in slightly better fits to the data than the regularized solution.

The remainder of the paper proceeds as follows: In section 3.2 we discretize the disordered kinetic model for numerical manipulation. In section 3.3 we calculate the model inverse. This basic inversion easily and rapidly identifies the appropriate multi-pool solution without prior assumption regarding the number and type of pools. Later in that section we show that the best fitting multi-pool solution is sensitive to noise in the data as both the number and type of best fitting pools fluctuate with the noise levels found in the data. Thus, in section 3.4 we address the noise sensitivity by employing a method of regularization to invert the data and provide details of its implementation. Finally, in section 3.5, we analyze litter decay data by assuming rates are distributed lognormally and fit just the two lognormal parameters to the data. Matlab codes for all numerical procedures and approaches are provided online in the Supplementary Material.

3.2 The Parallel Model of Decay

Because organic matter is composed of many different compounds, some are more resistant to degradation and break down more slowly than others [63, 64, 1, 34, 65]. Substrate heterogeneity suggests that degradation proceeds at different rates in parallel. This heterogeneity can result in either a discrete or a continuous distribution of decay rates.

We identify the best fitting rate distribution by inverting a continuous parallel model of decay. In this model, the fraction $g(t)$ of original mass remaining is described by a continuous superposition of exponential decays [49],

$$g(t) = \int_0^\infty p(k)e^{-kt}\mathrm{d}k \tag{3.1}$$

where $p(k)$ is the probability distribution of components with rate k and $\int_0^\infty p(k)\mathrm{d}k = 1$. We have previously found that litter decay rate distributions are well characterized by a lognormal $p(k)$, although other possible forms have been hypothesized [49, 18, 50, 41].

Mathematically, equation (3.1) is the Laplace transform of $p(k)$. We obtain the distribution $p(k)$ by taking the inverse Laplace transform of the data $g(t)$ [13]. To do this, we transform the model to facilitate calculations. We first make a change of variables from k to $\ln k$ as we expect to find a wide range of decay rates. Because probability is conserved, $p(k)\mathrm{d}k = \rho(\ln k)\mathrm{d}\ln k$, and equation (3.1) becomes

$$g(t) = \int_{-\infty}^{\infty} \rho(\ln k)e^{-kt}\mathrm{d}\ln k \tag{3.2}$$

where $\rho(\ln k)$ is the probability distribution in $\ln k$ space.

To numerically calculate $\rho(\ln k)$ we discretize equation (3.2). Although many discretization methods can be used, such as quadrature [112, 124], we choose to discretize (3.2) at n nodes λ_j spaced uniformly along the $\ln k$ axis between the limits $\ln k_{\text{min}} < \ln k < \ln k_{\text{max}}$. The nodes are therefore spaced at intervals of width

$$\Delta\lambda = \frac{\ln k_{\text{max}} - \ln k_{\text{min}}}{n}.$$

This discretization is chosen in order to provide resolution over the appropriate wide range of k. [2] Equation (3.3) can then be written in matrix form as

$$\mathbf{g} = \mathbf{A}\vec{\rho}. \tag{3.3}$$

$\vec{\rho}$ is a vector of length n, representing the average value of $\rho(\ln k)$ over an interval $\Delta\lambda$, centered at λ_j,

$$\rho_j = \frac{1}{\Delta\lambda}\int_{\lambda_j - \frac{1}{2}\Delta\lambda}^{\lambda_j + \frac{1}{2}\Delta\lambda} \rho(\ln k) \mathrm{d}\ln k.$$

The $m \times n$ matrix $\mathbf{A}$ is the discrete Laplace transform operator with elements

$$A_{ij} = e^{e^{\lambda_j} t_i} \Delta\lambda,$$

and $\mathbf{g}$ is a vector of m predicted time series data points $g_i = g(t_i)$.

We proceed to identify the underlying rate distribution $\rho(\ln k)$ from an observed decay time series $\hat{g}(t)$ by inverting the model (3.3) and solving for $\vec{\rho}$.

3.3 Direct Inversion with Constraints

Our approach in this section directly calculates the inverse Laplace transform of data with non-negative constraints. This approach identifies a handful of separate and distinct pools which are present during decay, having values $\rho_j > 0$; the remaining pools are not present and have values $\rho_j = 0$. This technique therefore provides a direct estimate of the best fitting multi-pool model, expressed as a discrete distribution $\rho(\ln k)$. We find however that the multi-pool solution is very sensitive to noise in the data and is overparametrized. We therefore proceed in section 3.4 to refine this approach using a regularization technique to invert the data [13], providing a continuous rate distribution $\rho(\ln k)$ which is both simpler and less sensitive to noise.

3.3.1 Calculating the constrained direct inverse

The distribution $\vec{\rho}$ can be directly calculated by computing the inverse of (3.3) from the measured decay data $\hat{\mathbf{g}}$ which includes the first data point $g_1 = 1$ at $t_1 = 0$,

$$\vec{\rho} = \mathbf{A}^{-1}\hat{\mathbf{g}} \tag{3.4}$$

---

[2] While the domain of the inverse Laplace transform is infinite, we find that for litter decay data, the tails of the distributions vanish numerically and that the distribution is typically contained between $10^{-6}\mathrm{yr}^{-1} < \mathrm{k} < 10^{4}\mathrm{yr}^{-1}$. We therefore set the values $k_{\max} = 10^{4}\mathrm{yr}^{-1}$ and $k_{\min} = 10^{-6}\mathrm{yr}^{-1}$. We choose the number of discretization steps, n, to provide a reasonable resolution of $\rho(\ln k)$ without being computationally burdensome. $n \approx 100$ or $n = 16m$ was satisfactory.

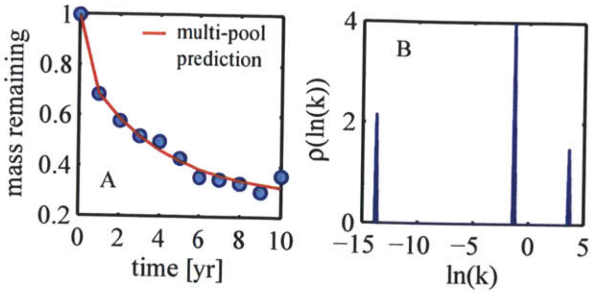

Figure 3-1: Inverse Laplace transform of a LIDET dataset. A: Decay data $\mathbf{g}$ of a LIDET dataset. Line is the fit to the data corresponding to the distribution $\rho(\ln k)$ shown in B. B: Distribution $\rho(\ln k)$ obtained by numerical inversion of the data $\hat{\mathbf{g}}$ using the constrained non-negative least squares algorithm.

Because solutions to (3.4) fit the noisy data exactly, negative values of ρ_j are possible. In order to find solutions with non-negative ρ_i, we instead solve the constrained least squares problem

$$\min_{\vec{\rho}} \|\mathbf{A}\vec{\rho} - \hat{\mathbf{g}}\|^2 \tag{3.5}$$

such that

$$\rho_j > 0 \tag{3.6}$$

$$\sum_{j=1}^{n} A_{1j}\rho_j = 1, \tag{3.7}$$

where $\|\mathbf{x}\| \equiv \sqrt{\sum x_i^2}$ is the norm of vector $\mathbf{x}$. The elements A_{1j} corresponds to $t_1 = 0$ and constraint (3.7) ensures that $\vec{\rho}$ is a probability distribution, which sums to one. We use the Matlab function `lsqnonneg.m`, which employs Lagrange multipliers [128] to calculate the non-negative solution to the least squares problem (F.4). Constraint (3.7) is met by weighing the first data point at $g(0) = 1$ more heavily than the others, although an additional Lagrange multiplier could be used. Because $m < n$, A is underdetermined and has rank m. In Matlab 2008a and older, the function `lsqnonneg.m` uses the algorithm `mldivide.m` [129] to calculate $\mathbf{A}^{-1}\hat{\mathbf{g}}$. This algorithm utilizes a rank-revealing QR factorization with column pivoting which calculates $\vec{\rho}$ only from the m most linearly independent orthogonal components of A. As a result, it returns at most m non-zero components in the vector $\vec{\rho}$. In newer versions of Matlab, `lsqnonneg.m` uses the pseudo-inverse, `pinv.m` [130] to calculate $\mathbf{A}^{-1}\hat{\mathbf{g}}$. The pseudo-inverse identifies the solution $\vec{\rho}$ with minimum norm. For our problem, we find that results from both the newer and older versions of `lsqnonneg.m` are the same.

We apply this method to data from the Long-Term Intersite Decomposition Experiment Team (LIDET) study [21, 24, 22, 25, 99]. The LIDET study monitored the decomposition of 27 different

Table 3.1: This table shows the wide and seemingly unpredictable variation in the types of pool models that best fit 191 LIDET datasets. Each model was obtained by solving the non-negative least squares problem (F.4) and analyzing the resulting $\vec{\rho}$ to determine whether the pools were active(decaying exponentially), inert, or rapid, as discussed in the text. The first column lists the model type, the second column indicates kinds of pools present in that model, the third column provides the formula for the model, the fourth column gives the total number of pools in the model, the fifth column indicates the total number of parameters in the model, and the sixth column indicates how many LIDET datasets are best fit by that model. The sizes of the constant, rapid, and active pools in the models are is c, r, and a respectively. The size of one of the pools is determined by the others since mass fractions must sum to one.

model type	type of pools	formula	pools	para-meters	# data-sets
1	active	e^{-kt}	1	1	33
2	active, inert	$(1-c)e^{-kt}+c$	2	2	30
3	active, rapid	$(1-r)e^{-kt}$	2	2	37
4	active, inert, rapid	$(1-r-c)e^{-kt}+c$	3	3	25
5	2 active	$ae^{-k_1t}+(1-a)e^{-k_2t}$	2	3	39
6	2 active, inert	$ae^{-k_1t}+(1-a-c)e^{-k_2t}+c$	3	4	14
7	2 active, rapid	$ae^{-k_1t}+(1-a-r)e^{-k_2t}$	3	4	10
8	2 active, inert, rapid	$ae^{-k_1t}+(1-a-r-c)e^{-k_2t}+c$	4	5	3

types of litter, including needles, leaves, roots, wood, grass, and wheat distributed amongst 28 different locations across North America ranging from Alaskan tundra to Panamanian rainforests. Litter was collected and then re-distributed in litter bags at different sites in order to investigate the effect of composition, ecosystem, and climatic parameters on decomposition. Litter bags were collected and analyzed each year for up to ten years, with four replicates for each site, litter type, and removal time. Bags at tropical and sub-tropical sites were more frequently collected at three to six month intervals.

An example of a decay dataset is shown in Figure 3-1A. The rate distribution ρ corresponding to the solution of (F.4 - 3.7) is shown in Figure 3-1B. Three pools are associated with the decay shown in Figure 3-1C; a very rapid pool, a moderately labile pool, and an extremely slow pool. However, by varying the search domain of decay rates k_{min} and k_{max} we find that the slow pool always takes the smallest value, k_{min}, which in this case is 10^{-6}yr. This suggests that the pool at k_{min} represents an inert, or constant mass fraction. The fast pool on the other hand is located at $k = 38.5\text{yr}^{-1}(1.4\text{wk}^{-1})$. The amount remaining of this fast pool at the first measurement $t = 1$ yr is $e^{-38.5} = 1.9\text{E}-17$ which is just past the double precision limit of 16 significant figures. Therefore this pool is numerically zero for every measured data point, and is therefore indistinguishable from an instantaneous decay. The inversion therefore suggests that this dataset contains three types of pools: a rapid, instantaneous pool; an exponentially decaying, active pool; and a constant, inert pool. Because the inert and instantaneous pools do not have rates associated with them, this dataset is described by three parameters.

Models with various combinations of active, inert and instantaneous pools have been previously suggested to describe litter decay [24,73]. In 1945, Henin and Dupuis [73] suggested using an inert pool to represent the transformation of incoming carbon to stabilized soil carbon. However, because the turnover time is the mean inverse rate $\langle k^{-1} \rangle$ of each individual pool [41,13], decomposition models containing an inert pool have an infinite turnover time and result in the unphysical situation where soil organic matter grows indefinitely. These models require additional parameters and heuristics to calculate an effective rate of turnover [25,24].

The direct inversion technique provides an approach for identifying the simplest multi-pool solution which best fits the data. It generally provides solutions consistent with fitting different multi-pool models which contain various types of pools and choosing the model having the least error. This approach therefore gives a simple, direct estimate of the appropriate multi-pool solution and answers the question "how many pools?". The multi-pool approach however exhibits some serious shortcomings, which we discuss in the following sections.

3.3.2 Analysis of the multi-pool solutions

We use the same inversion technique to find the best multi-pool solution associated with 191 datasets chosen [13] from the LIDET study. We then determine whether pools are active, inert, or rapid as described above. If any pool contains a mass fraction of less than 1%, we neglect that pool and distribute its mass proportionately amongst the remaining pools, as these small pools require additional model parametrization but have a negligible effect on the residual error.[3] Table 1 shows the results of the inversion and indicates that eight different types of multi-pool models are needed to describe all 191 of these LIDET data sets. 17% of the datasets are described by one pool, 56% are described by two pools, and 27% are described by three or more pools. The maximum number of pools found was four and all datasets are characterized by 5 or less parameters. The data from Figure 3-1A is of type 4 in Table 1; 24 other LIDET datasets also exhibit the same behavior. This information can be used to determine some patterns underlying litter decomposition. For example, we find that 29% of all datasets involving either needles or wood appear to decay exponentially while roots on the other hand decay exponentially only for 2% of their datasets. Roots tend to have 3 or 4 pools for 48% of their datasets, while leaves, needles, and wood each have 3 or 4 pools less than 25% of the time. However, extracting further information from the inversion results in order to determine which model and parameters are appropriate for different combinations of tissue types, substrates, or environments is difficult if not impossible since there is too much variation in $\vec{\rho}$ and model type. Indeed this is one of the shortcomings of the multi-pool model. It is too overparametrized to be useful in identifying fundamental constitutive relations between composition and dynamics.

---

[3] This occurred for only 7/191 datasets, 5 of which were two pool models that became one pool models.

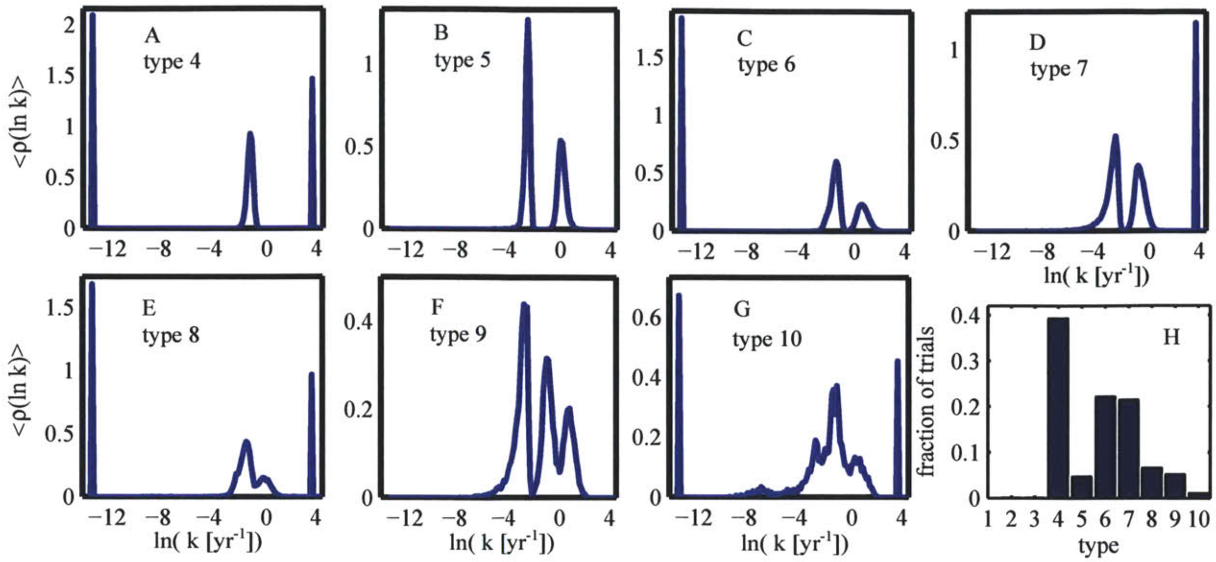

Figure 3-2: Distributions $\rho(\ln k)$ for 60000 trials of adding random noise to the predicted decay in Figure 1A. For each trial, $\rho(\ln k)$ is estimated using the method discussed in section 3.3.1. The pools in each solution are then analyzed to determine the model type as described in the text and Table 1, except for model type 9 which is a three active-pool model and type 10 which is any combination of pools not described by models 1-9. Figures A-G shows the mean $\langle\rho(k)\rangle$ for each model type, calculated by summing all of the observed distributions for each type and dividing by the total number of times each type was observed. H: Fraction of trials that each model type was identified.

3.3.3 Sensitivity to noise

In this section we test the direct inversion method and evaluate the appropriateness of multi-pool models for describing noisy decay data. The standard shortcoming of a multi-pool solution is its inherent sensitivity to noise in the data $\hat{\mathbf{g}}$ [113, 118], as models with differing numbers of pools can well approximate one another to a high degree of accuracy [121]. We proceed to investigate the solution's sensitivity to noise levels present in litter decay data.

If we assume that the multi-pool model is correct, then the line in Figure 3-1A is the true decay $g(t)$. We then estimate the variance of the data noise σ_ϵ^2 by calculating its maximum likelihood $\sigma_\epsilon^2 = \sum_i \hat{g}_i{}^2/m$ [131], assuming the noise is Gaussian with zero mean. We then generate numerical trials of decay data by adding Gaussian noise of the same strength, in this case $\sigma_\epsilon = 0.024$, to the predicted decay $g(t)$. Then, for each trial we recalculate the associated inverse solution $\vec{\rho}$. The results of the inversion of 60000 numerical trials is shown in Figures 3-2A-G. There is wide variation in the active pools just due to the noise in the data. Figure 3-2B shows how the model type varies due to noise. The "true" model, type 4, is only identified 39% of the time. The other LIDET datasets show similar results. These results highlight why it is difficult to gain fundamental understanding of decomposition processes from fitting a multi-pool model to the data.

3.4 Regularized Inversion Method

Because the inverse Laplace transform (3.4) is sensitive to noise in the data, it is ill-posed [115,124]. In the previous section, we have demonstrated that the inverse Laplace transform with non-negative constraints is also ill-posed [121] at the level of noise present in our decay functions $g(t)$. There are a number of approaches to solve the inverse Laplace transform robustly, without being sensitive to small changes in the data. These include Laplace-Padé approximation [121], the phase function method [132], regularization [124,115,133,116], and others [133]. We choose regularization because of its straightforward implementation. Regularization is commonly applied elsewhere, for example in the analysis of NMR spin relaxation [112] in biological tissue [113] and in porous media [117,118].

Like most other robust inverse Laplace transform methods, regularization works by determining the solution that, in principle, fits the data without fitting the noise. This solution represents a trade off between simplicity and accuracy. Here we choose a specific type of regularization called Tikhonov regularization [124, 115, 116] to calculate a solution to the constrained inverse problem (Eqs. F.4-3.7) because this method well handles the constraints on $\rho(\ln k)$. The goal of Tikhonov regularization is to minimize both the residual error and the complexity of the solution $\rho(\ln k)$. Solution complexity is assumed to be associated with the roughness, or the intensity of fluctuations, in $\rho(\ln k)$. Here, we measure roughness by the norm of the first derivative of the solution vector,

$$\left\|\frac{\mathrm{d}\rho(\ln k)}{\mathrm{d}\ln k}\right\| = \left(\sum_i \left(\frac{\rho_{i+1} - \rho_i}{\ln k_{i+1} - \ln k_i}\right)^2\right)^{1/2} = \|\mathbf{R}\vec{\rho}\|, \tag{3.8}$$

where $\mathbf{R}$ is the bi-diagonal first derivative operator, with an additional first row $[1\ \mathbf{0}]$ and additional final row $[\mathbf{0}\ \ -1]$ to account for $\vec{\rho}$ being zero outside the domain $k_{\text{min}} < k < k_{\text{max}}$.

The regularization method proceeds by finding the solution $\rho(\ln k)$ that minimizes a sum of the residual error and the roughness:

$$\min_{\vec{\rho}} \|\hat{\mathbf{g}} - \mathbf{A}\vec{\rho}\|^2 + \omega\|\mathbf{R}\vec{\rho}\|^2, \tag{3.9}$$

where ω is the regularization parameter which controls the relative weight of the solution roughness to the residual error. Because it is unclear how much to weigh the roughness, ω is typically varied many orders of magnitude from $\omega \ll 1$, which emphasizes the importance of residual error, to $\omega \gg 1$, which emphasizes the smoothest solution. The best $\rho(\ln k)$ is determined by identifying the value of ω that sets an optimal balance between the residual error and the roughness.

3.4.1 The L-Curve

An approach for finding the optimal ω is to use the "L-curve" technique [115, 124]. An L-curve is generated by parametrically varying ω and solving (3.9) for ρ, obtaining values for the residual error norm $\|\hat{\mathbf{g}} - \mathbf{A}\vec{\rho}\|$ and roughness norm $\|\mathbf{R}\vec{\rho}\|$ for each value of ω. The L-curve is made by

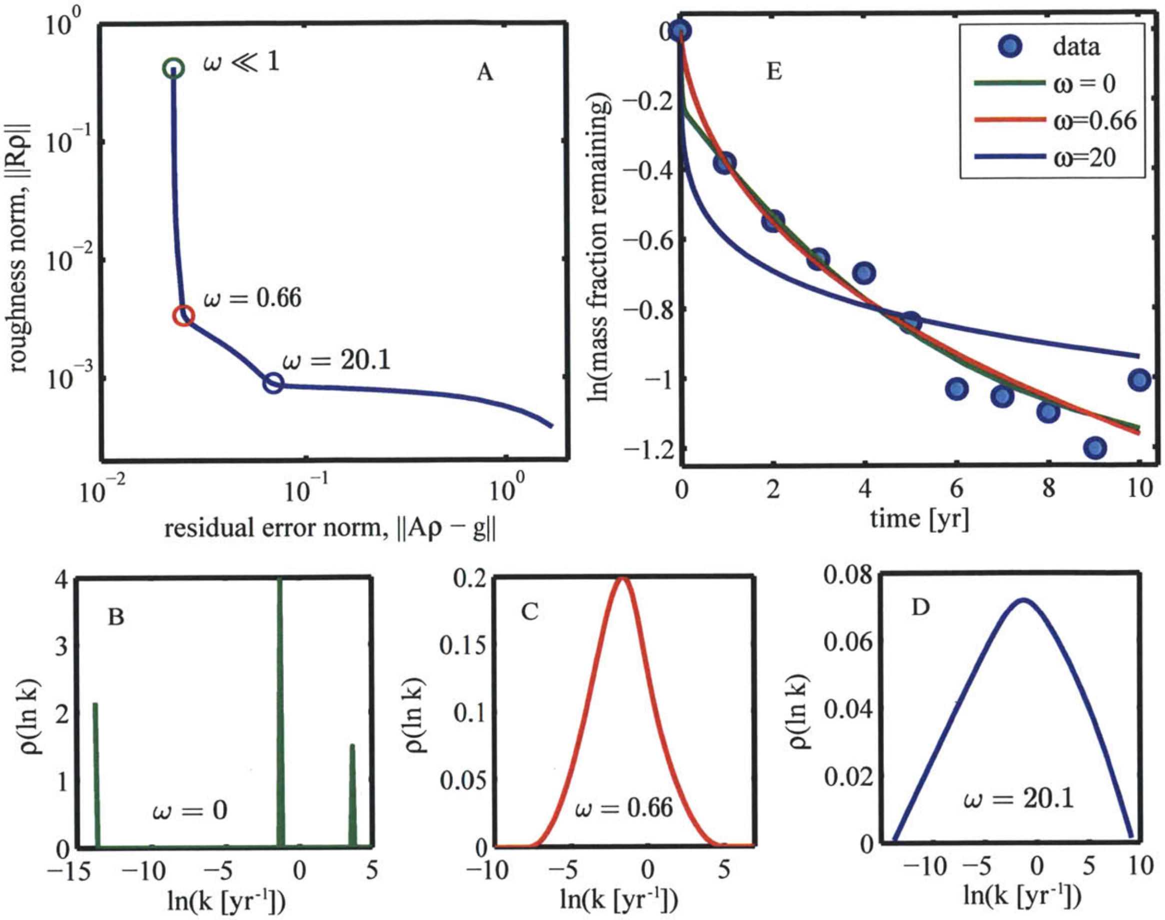

Figure 3-3: Determining the optimal value of regularization parameter ω via the "L-Curve" method [124, 115]. Counterclockwise from the upper left: A) The "L-curve", $\|R\hat{\rho}\|$ vs. $\|\hat{g} - A\hat{\rho}\|$. B) The unregularized solution, similar to the solution found at the top of the curve, where the regularization parameter $\omega \ll 1$. C) The solution associated with the corner of the curve corresponding to $\omega = .66$. D) The solution at the corner of the curve corresponding to $\omega = 20.1$. E) The data $g(t)$ as well as the decay predicted from the solutions in B,C, and D. The corner with $\omega = .66$ is chosen as the optimal solution since it predicts the data with significantly less residual error and has no trend in the residual, unlike the corner having $\omega = 20.1$ The optimal solution is associated with a corner of the curve.

plotting $\|\mathbf{R}\vec{\rho}\|$ vs. $\|\hat{\mathbf{g}} - \mathbf{A}\vec{\rho}\|$ on a log-scale, resulting in a characteristic "L" shape. Figure 3-3A shows the L-curve for the dataset shown in Figure 3-1A.

While each point on the curve is associated with an optimal solution for a certain value of ω, the corner of the L-curve is associated with the regularized solution. The corner represents the point where increasing the information contained in the solution no longer improves the residual error. Solutions at the upper right of the L-curve contain high information as n elements ρ_j are free and independent [134] whereas fewer modes are active in the smooth solution at the corner. The corner is also the point where relatively little decrease in information results in large increases

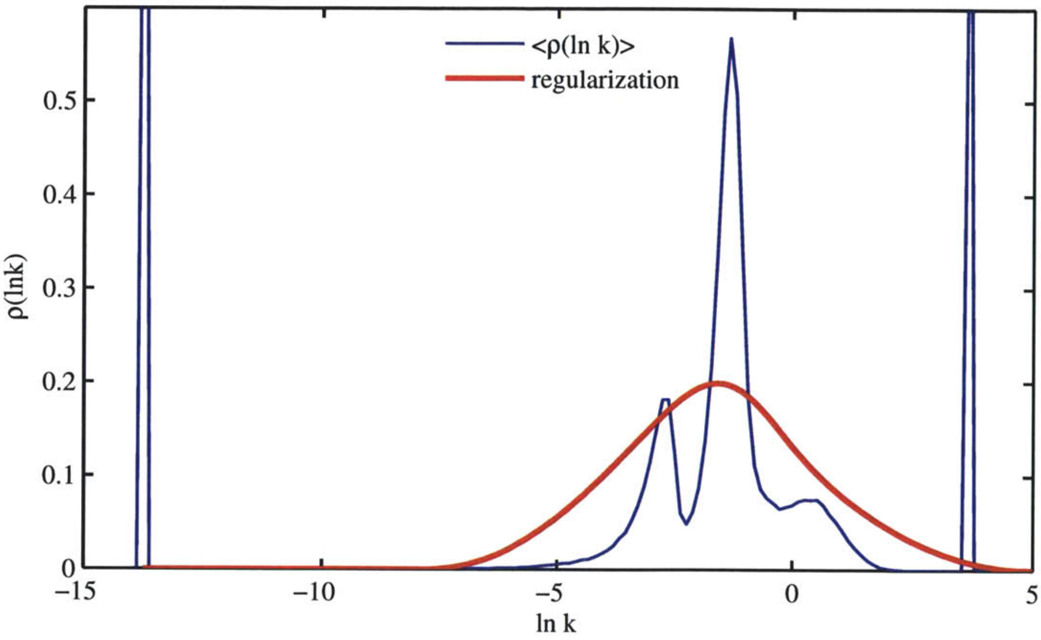

Figure 3-4: The average (blue) of all 60,000 noise-generated multi-pool solutions from Figure 3-2 plotted alongside the regularized solution (red). $\langle \rho(\ln k) \rangle$ is calculated by summing $\vec{\rho}$ from all 60,000 noise-generated trials and dividing the sum by 60,000.

in the residual error. Often, the corner is associated with a change in the number of maxima in $\rho(\ln k)$; solutions above the corner may have more modes than solutions below.

Generally, the L-curve method identifies a smooth solution which predicts a decay having a residual error approximately equal to the noise in the dataset. Solutions for three different values of ω are shown in Figures 3-3B-D with the unregularized solution shown in Fig. 3-3B. Decays predicted from these three solutions are shown in Fig. 3-3E.

When there are multiple corners, as shown here, there are a few approaches to choosing the appropriate one. We use the following method to identify the corner associated with the simplest solution:

1. For each value of ω, calculate the rank correlation [123] of the residual error and the corresponding P value.
2. Identify the region of the L-curve where the rank correlation of solutions have $P > 0.1$, suggesting no significant trend in the residual.
3. Choose the corner within the region of the L-curve having $P > 0.1$;
4. If a corner seems too rounded or undefined, the L-curve can also be plotted on linear $\|\mathbf{A}\vec{\rho} - \hat{\mathbf{g}}\|$ and $\|\mathbf{R}\vec{\rho}\|$ axes. The corner collapses onto a much tighter volume in linear space and can sometimes be more readily chosen if the corner in log space has large curvature.

5. If there are multiple corners with $P > 0.1$, compare more closely the residual trend in all corners, as values $P > 0.1$ do not necessarily indicate a lack of trend in the residual. If there is a noticeable difference in the quality of the fit associated with each corner, choose the corner that does not appear to have a trend in the residual. An additional test is to look at the error bias. Here, the error bias is simply the sum of residual error divided by the error norm, $\frac{|\sum(\mathbf{A}\vec{\rho}-\hat{\mathbf{g}})|}{||\mathbf{A}\vec{\rho}-\hat{\mathbf{g}}||}$. Experience with this data suggests choosing solutions with error bias $< .25$. If multiple corners still have a bias $< .25$ and a similar strength in trend in their residual error, choose the corner with lowest roughness. Usually the corner having lower residual error and higher roughness tends to be not biased and to have no trend in the residual.

We apply this approach to choose the appropriate corner in Figure 3-3A. The distribution corresponding to the lower corner (Figure 3-3D) predicts a decay that has a large error compared to the noise in the data and has a significant trend in the residual error as seen in Figure 3-3E. However, the distribution corresponding to the higher corner (Figure 3-3C) predicts a decay which is hardly distinguishable from the decay predicted by the unregularized solution, shown in green, after $t = 1$. Therefore, the distribution shown in Figure 3-3C is considered the optimal, regularized solution.

3.4.2 Results of regularization

Figure 3-4 shows the overall variation in the inversions due to noise by summing all 60,000 multi-pool solutions from Figures 3-2A-G. We also plot the solution $\rho(\ln k)$ from the regularized inversion. The regularization provides a simpler solution that captures the variation of the rate heterogeneity associated with uncertainty of the multi-pool solutions.

More specifically, we find that this heterogeneity takes a certain form: the regularized solution presented in Figure 3-3C appears lognormal, i.e. Gaussian in $\ln k$ space. We have shown in a separate publication [13] that the entire LIDET dataset is lognormal on average. This lognormal description is considerably simpler than the multi-pool approach as the rate heterogeneity and all dynamic information are now contained within just two parameters, a mean μ and a variance σ^2 of the order of magnitude of decay rates. This approach also emphasizes the view that the quality of litter and young soil organic matter quality is continuously distributed.

3.5 Fitting Lognormal Parameters to Decay Data

The ubiquity of the the lognormal distribution suggests that litter decay can be well described by the two lognormal parameters, μ and σ. In this section, we assume a model of lognormally distributed rates and present a procedure to estimate the values of μ and σ that best fit the decay data. We then compare its results to the results of the regularized inversion.

We proceed by substituting the Gaussian distribution,

$$\rho(\ln k; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\ln k-\mu)^2/2\sigma^2}, \tag{3.10}$$

for $\rho(\ln k)$ in equation (3.2), providing a prediction of the mass fraction $g(t)$ remaining as a function of the parameters μ and σ:

$$g(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-(\ln k-\mu)^2/2\sigma^2} e^{-kt} \mathrm{d} \ln k. \tag{3.11}$$

We then identify the values of μ and σ which best fit the data by solving the non-linear least squares minimization problem

$$\min_{\mu,\sigma} \sum_{i=1}^{n} \left(\hat{g}_i - \int_{-\infty}^{\infty} \rho(\ln k; \mu, \sigma) e^{-kt} \mathrm{d} \ln k \right)^2, \tag{3.12}$$

where $\hat{g}_i$ are the measured data points. We solve (3.12) using Matlab's non-linear least squares solver `nlinfit.m`. This is repeated for all 191 LIDET datasets that are appropriately described by a superposition of exponential decays [13].

Figure 3-5 shows a comparison of the residual error associated with the three different approaches: direct inversion (F.4-3.7), regularized inversion (3.9), and the two-parameter fit (3.12). The root mean square error (RMSE) of the fit to the data is calculated for each dataset and the cumulative density function of the RMSE rescaled by the total number of datasets (191) is plotted; the vertical axis shows the number of datasets having an RMSE smaller than the value on the horizontal-axis. The RMSE of the unregularized inversion technique (green) is clearly smaller than that of the other two techniques, but these solutions suffer from the problems discussed in section 3.3.3. Mean RMSE values for all 191 datasets are 0.050, 0.056, and 0.055 for the direct inversion, regularized inversion, and lognormal model respectively. Surprisingly, solutions from the two-parameter fit (magenta) appear to have slightly smaller residual error than solutions identified from the regularization technique (blue). Thirty-two of the datasets are predicted by both the two-parameter fit and regularization method to decay exponentially; the two-parameter lognormal model however predicts the decay from 126 of the remaining 159 datasets better than the distribution obtained by regularization. Therefore, the lognormal model fits the data equally or better than the regularized solution for 158/191 or 83% of the datasets. This is unexpected because the lognormal model has only two degrees of freedom. The marginally poorer fit of the regularization method is due to the emphasis of the method on smoother (wider) solutions. We proceed to investigate this effect by comparing the fitted parameters σ to the standard deviation of the regularized inversions $\rho(\ln k)$.

Figure 3-6 compares the fitted values μ and σ to the mean $\langle \ln k \rangle$ and standard deviation $\sqrt{\langle (\ln k - \langle \ln k \rangle)^2 \rangle}$ associated with the regularized solutions $\vec{\rho}$. Figure 3-6A shows that values of μ and the mean of the inversion $\rho(\ln k)$ are consistent with one another. There is a departure

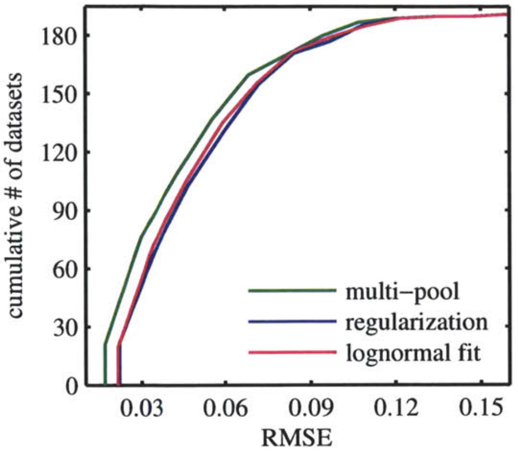

Figure 3-5: Comparison of root-mean-square error (RMSE) associated with the three different parameter estimation procedures. For each one, the RMSE of the predicted decay is calculated for 191 LIDET data sets. The y-axis shows the number of datasets having an RMSE smaller than the value of RMSE on the x-axis. Shown are the RMSE of the unregularized multi-pool solution (green), regularized solution (blue), and two-parameter nonlinear fit (magenta).

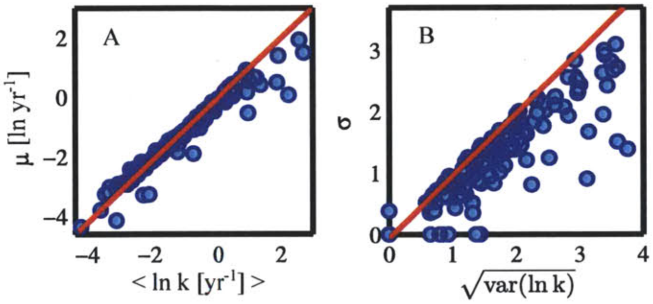

Figure 3-6: Comparison of μ and σ estimated from the nonlinear fitting procedure to the mean $\langle \ln k \rangle$ and standard deviation of the regularized distribution $\vec{\rho}$. Figure A: the mean $\langle \ln k \rangle$ of $\rho(\ln k)$ vs. μ. B: the standard deviation of $\rho(\ln k)$ vs σ.

at high $\langle \ln k \rangle$ because some of these inversions are bi-modal: one mode is active whereas the other mode is distributed over extremely fast rates associated with instantaneous decays. These bi-modal datasets are effectively described by an instantaneous mass loss followed by decay that proceeds with a distribution of rates. Figure 3-6B, on the other hand, tells another story. Values of σ tend to be less than the standard deviation of $\rho(\ln k)$, indicating that the regularized solution is wider than the best fitting lognormal distribution. This is a consequence of weighing the solution roughness during regularization. A narrower lognormal solution with smaller residual error exists, but the weight of the roughness is large enough that the regularization method chooses a slightly wider solution which fits the data almost as well. Collectively, Figures 3-5 and 3-6B indicate that when rates are heterogeneous, the transition from unimodal to multimodal solutions near the corner of the L-curve occurs before finding the unimodal solution with the smallest residual error.

These results indicate that the regularization method is useful for identifying that general solution shapes and trends. If regularization suggests that rate distributions are lognormal, then fitting μ and σ to the data identifies more precisely the specific lognormal distribution that best fits the data.

3.6 Conclusions

Direct calculation of the inverse Laplace transform with a non-negativity constraint provides the best fitting multi-pool solution without specification of the number of pools a priori. However, this multi pool solution is very sensitive to small changes in the decay function $g(t)$ [121] as the number of pools and the rates of each pool vary widely due to the level of noise associated with litter bag data. The regularization method described here is robust to noise and commonly indicates that a lognormal distribution provides a concise representation of the rates associated with decay data. If the lognormal distribution may be assumed, a marginally better fit to the data is found by directly estimating the lognormal parameters μ and σ from the decay $g(t)$.

Matlab codes for all numerical procedures (direct inversion, multi-pool estimation, regularized inversion, and the two-parameter fitting procedure) are provided online at http://www.biogeosciences-discuss.net/9/3795/2012/bgd-9-3795-2012-supplement.zip.

Chapter 4

Carbon Transit through Degradation Networks

Abstract

The decay of organic matter in natural ecosystems is controlled by a network of biologically, physically, and chemically driven processes. Decomposing organic matter is often described as a continuum which transforms and degrades over a wide range of rates, but it is difficult to quantify this heterogeneity in models. Most models of carbon degradation consider a network of only a few organic matter states which transform homogeneously at a single rate. These models may fail to capture the range of residence times of carbon in a decay network. Here we assume that organic matter is distributed amongst a continuous network of states which transform with stochastic, heterogeneous kinetics. We pose and solve an inverse problem in order to identify the rates of carbon exiting the underlying degradation network and apply this approach to plant matter decay throughout North America. This approach provides the complete information regarding carbon retention in a decay system without knowing the details of state transformations. We find that the exit rates are approximately lognormal, suggesting that carbon flow though a complex degradation network can be described with just a few parameters. This result also indicates that the serial and feedback processes in a degradation network can be well approximated by a parallel process model.

4.1 Introduction

The retention of particles in networks is associated with problems in ecosystems modeling [45,46,93,92], drug kinetics [88,89,135], logistics [136], industrial processing [137], and other fields [138]. Flow diagrams [137], Markov models [89], and compartmental models [138] are used to model these networks, where information flows from one compartment to another in parallel, series or with feedback. Often, kinetic parameters of these systems are unknown. These parameters can be identified by fitting model outputs to dynamic data. However, identifying system parameters when the underlying network structure is unknown remains a challenge.

Here, we investigate the retention of carbon in terrestrial organic matter. Plant litter is initially comprised of many organic components which degrade at different rates [63, 64, 1, 34, 65]. As decomposition proceeds however, organic matter is transformed. Extracellular enzymes break down polymers into shorter chains, acids break free, and decomposers transform compounds into decomposer biomass and other organic byproducts of the decomposition process. These compounds may then interact chemically [32,66,1], forming humic and other hard-to-degrade compounds [65,67,47]. Furthermore, particulate and dissolved carbon bond and sorb to clays and minerals [69–73] forming organo-mineral complexes that also affect decomposability. In this manner, organic carbon is transformed from one pool to another by means of both serial and feedback processes. Most models of soil organic matter decomposition include these serial and feedback loops between various pools [45, 46, 92, 93], but the network architecture is based purely on intuition about decomposer food chains and empirical results regarding the precursors for more recalcitrant soil organic matter. Other models treat soil organic matter as a continuum, where transformations occur amongst a continuum of states [10, 15]. It remains a challenge however to characterize these systems.

Previously [13] we utilized a continuous model of decay heterogeneity to analyze decay data from the Long-term Inter-site Decomposition Experiment, (LIDET) [99, 21, 22, 24, 23, 25, 20]. In that model, organic matter is assumed to be composed of heterogeneous particles which decay with a distribution of first-order decay rate constants k. The fraction $g(t)$ of organic mass remaining is a superposition of exponential decays,

$$g(t) = \int_0^\infty p(k) e^{-kt} dk, \tag{4.1}$$

where $p(k)$ is a probability distribution having the properties $p(k) > 0$ and $\int_0^\infty p(k) = 1$. By inverting equation (4.1), we found that $p(k)$ is on average lognormal. Therefore, organic matter decomposition from the LIDET study can be described by a two-parameter lognormal continuum of parallel decay rates.

Here we show that equation (4.1) can also describe decomposition proceeding through a complex network of transformation and decay mechanisms. In this formulation, $p(k)$ represents the distribution of Poisson rates at which particles exit from all pathways in the network. By estimating $p(k)$ we capture the mass dynamics of the system without needing to model the details of transformations across the entire network. We show that, for the case of the LIDET study, the same lognormal decay rate distribution well approximates the rates of exit from a stochastic network of transformation processes. Interpreting the network as $p(k)$ provides a unique way to visualize and pinpoint key dynamic information regarding carbon in the system and reveals the timescales of the retention of organic carbon in decomposition systems.

The work presented here relates to previous efforts to simplify degradation models by eigenvalue decomposition and diagonalization [18]. Bolker et al. diagonalized small decomposition models such as the CENTURY model, which assume organic matter transforms between only a few different states. They calculated system eigenvalues (diagonalized decay rates) and interpreted the meanings of positive and negative weights (or loadings) of the diagonalized exponential decays. The CENTURY model was simplified by reducing the number of parallel compartments and parameters needed to equivalently describe organic matter decay, and they suggested further simplifications may be obtained by assuming a distribution of parallel rates. We build upon that work by assuming Here however we consider transformations between states that take place heterogeneously at a distribution of rates. We build upon the previous work by use the diagonalized parallel approach to describe the heterogeneous organic matter transformation continuum and provide an inverse method for determining its shape. Using this method to analyze plant matter decay data, we find that the decay continuum itself is close to diagonal. This result supports assumptions used in the popular decompostition models studied by Bolker et. al.

In section 4.2, we introduce decomposition in terms of network of heterogeneous reactions. We present a simple two state network as a simple illustration of the methods used throughout the paper. In section 4.4, we explore examples of compartmental decomposition systems where mass exchange between compartments takes place at a distribution of rates. We investigate the properties of heterogeneous compartments in series and in feedback. In section 4.5 we investigate the behavior of the CENTURY terrestrial respiration model [45] with compartments that have heterogeneous kinetics. We find that if rates of exchange between pools are widely dispersed, the compartmental structure of the system is effectively lost. As a result, a heterogeneous CENTURY model appears simpler than the original discrete version with homogeneous compartments. In section 4.6, we consider the case of a more general stochastic system composed of a continuum of plant litter, microbial, and soil organic matter states. In section 4.7, we calculate the distribution of rates associated with the LIDET datasets constrained so that these rates represent the exit rates from a transformation network. Finally, in section 4.8, we find that the distribution of exit rates from the network are approximately lognormal, and can be approximated by a parallel system.

Notations

Notations for the variables used here are as follows:
t = time
G = total system mass
G_0 = initial system mass
g = fraction of initial mass remaining in the system, $g = G/G_0$.
$\mathbf{x}$ = vector of all state masses
$x_{i,j}$ = mass of state j in pool i
n = number of states per pool
$k_{i,j}$ = rate of transformation out of state j in pool i
$p_{i,j}$ = mass fraction entering pool i partitioned into state j.
$f_{i,j,l}$ = conversion fraction. Fraction of mass transferred from state j in pool i, to pool l.

$1 - \sum_j f_{i,j,l}$ = fraction of mass going to the exit or absorbing state: inorganic carbon or leached carbon.
y = total number of pools in the network.
n = number of states in a pool.
z = total number of states in the network.
$\lambda_i = i^{th}$ eigenvalue of the decay system.
$\Lambda = z \times z$ matrix with diagonals λ_i.
$\mathbf{u}_i = i^{th}$ eigenvector of the decay system.
$\mathbf{U} = z \times z$ matrix having columns $\mathbf{u}_i$.
$\mathbf{S} = z \times z$ diagonal matrix with i^{th} diagonal $= \sum_j u_{i,j}$
r_i = initial mass fractionation associated with the i^{th} eigenvalue.
$r(k)$ = initial mass fractionation onto the eigenvalues k of a decay system with a continuum of states. Continuous distribution of exit rates k. Exit rate function.
$v(\ln k)$ = continuum of exit rates in log rate space. Log exit rate function.

4.2 Degradation as a Network of Transformations

We consider decomposition as a network of reactions which transform organic carbon between a widely dispersed network of states. Transformations to and from each state occur at many rates depending on physical and chemical conditions, decomposer community and other stochastic variables. In other words, transformations affect both bio-availability, whether appropriate enzymes that can cleave the substrate are available [74], and bio-accessibility, whether an appropriate decomposer can physically access the substrate [74]. These transformations likely occur amongst a continuum of soil organic matter states [38, 14, 10, 15]. With this in mind, we suggest that organic matter transforms between a distribution of states, at a distribution of rates.

We proceed by quantitatively tracking a carbon atom on its journey through various transformations in a decomposition network. There are two equivalent ways of describing the movement and transformation of mass amongst different states. One method is to use a deterministic model which describes the multi-state system by a set of first order linear differential equations. The other method is to consider a stochastic model, where particles have a given probability of transforming from one state to another over a small time period. The two approaches are equivalent for the purpose of describing overall mass dynamics of degradation systems. Both types of models have been used to model decomposition but often contain a small finite number of states. The stochastic approach has been implemented via an absorbing Markov chain [94]. The deterministic approach is more common, usually implemented via compartmental models such as CENTURY [45], Roth-C [93], the carbon turnover model of Van Veen and Paul [44], and many others [9]. These compartmental models are equivalent to continuous-time Markov processes. In all of these models,

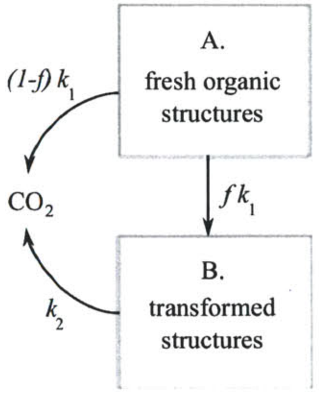

Figure 4-1: A two-state network. State A may be a fresh litter structure, while state B is a transformed organic structure. Particles leave the first state at a first-order rate k_1. These particles transfer to the second state with a probability f and therefore transform at the rate fk_1. Particles transfer from the first state to the exit state, CO$_2$, at the rate $(1-f)k_1$. In this simple model, particles transform from the second state only to the exit state, and do so at the rate k_2.

transformations between states occur continuously in time at the rates k. The main characteristic of Markov processes is transformations of particles from one state to another are memory-less (depend on only the current state). In this paper we often use the deterministic formulation to calculate dynamic properties of the system. However it is helpful to keep the stochastic Markov interpretation in mind when thinking of organic carbon transformations. We provide stochastic interpretations in the text when relevant.

4.2.1 Two-state example

Figure 4-1 shows a pedagogical device where organic matter exists in just two states: an initial state and a transformed state. We assume transformations from the initial state occur at the first-order rate constant k_1, with probability f of transforming to the second state, and probability $(1-f)$ of transforming to the exit state (carbon dioxide). Carbon in the second state then exits at the rate k_2. The dynamics of this simple system follow equations (4.2) and (4.3), where x_i is the concentration or amount of carbon in state i.

$$\frac{\mathrm{d}x_1}{\mathrm{d}t} = -k_1 x_1 \tag{4.2}$$

$$\frac{\mathrm{d}x_2}{\mathrm{d}t} = -k_2 x_2 + f k_1 x_1 \tag{4.3}$$

Equations (4.2) and (4.3) can be more compactly written in matrix form as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x} \tag{4.4}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} -k_1 & 0 \\ fk_1 & -k_2 \end{bmatrix} \tag{4.5}$$

having the initial condition $\mathbf{x}(0)$.

The solution $\mathbf{x}(t)$ is derived in Appendix 4.10.1 and is a superposition of two exponential decays,

$$\mathbf{x} = \alpha_1 \mathbf{u_1} \mathrm{e}^{-\lambda_1 \mathrm{t}} + \alpha_2 \mathbf{u_2} \mathrm{e}^{-\lambda_2 \mathrm{t}}, \tag{4.6}$$

where $\mathbf{u}_i$ are the eigenvectors of the system and α_i are the weights for each exponential decay.

Substituting the initial condition $\mathbf{x}(0)$ into (4.6) provides the weights α_i. The total mass of the system, $G(t) = x_1(t) + x_2(t)$ therefore decays as

$$G = r_1 \mathrm{e}^{-\lambda_1 \mathrm{t}} + \mathrm{r}_2 \mathrm{e}^{-\lambda_2 \mathrm{t}} \tag{4.7}$$

where r_i is determined by summing the states x_i of (4.6)

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} \sum_i u_{1,i} & 0 \\ 0 & \sum_i u_{2,i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}. \tag{4.8}$$

$u_{1,i}$ are the components i of the eigenvector $\mathbf{u_1}$ and $u_{2,i}$ are the components i of the eigenvector $\mathbf{u_2}$, $i = 1, 2$. For the system (4.5), the values of r_i are

$$r_1 = x_1(0)\left(1 + \frac{fk_1}{k_2 - k_1}\right), \tag{4.9}$$

$$r_2 = x_2(0) - x_1(0)\frac{fk_1}{k_2 - k_1}. \tag{4.10}$$

The lesson of this elementary exercise is that although the system (4.2) contains two pools in series, its total mass decays as if it were two pools in parallel [39, 18]. The weights r_i of each exponential decay are related to both the eigenvectors of the system and the initial conditions. Furthermore, the rates of parallel decay reflect the rates of the underlying serial pools. These properties are consistent with all examples of decomposition networks throughout the paper.

4.2.2 A continuum of transformations

We now suppose that pools of carbon are composed of heterogeneous particles and that transformations between pools occur over a distribution of rates. We introduce this concept by fractionating each compartment in the simple model of Figure 4-1 into a distribution of n states. These particle-states (pseudo-compartments [88]) transform at random rates [139] as shown in Figure 4-2.

The heterogeneous transformations shown in Figure 4-2 may be represented by the set of $2n$

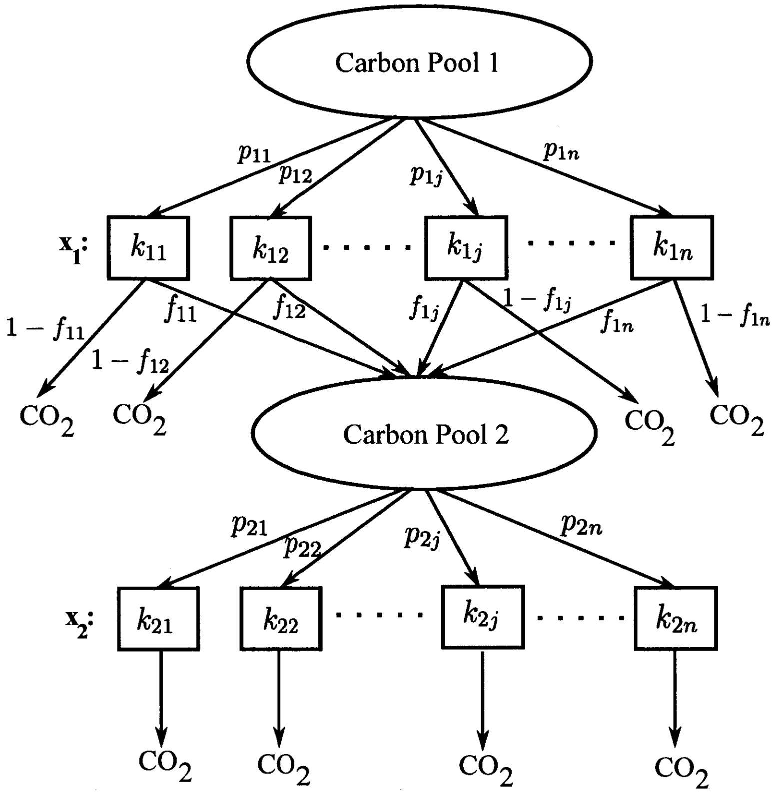

Figure 4-2: Serial processes with distributed rates. Transfers between pools of organic matter occur a distribution of rates. If the processes are biotic, carbon may be respired to CO2, and only a fraction of carbon is passed to the next pool. Carbon in the first pool is partitioned into states $\mathbf{x}_1$ by the fractionation vector $\mathbf{p}_1$, with $p_{1,i}$ determining the fraction of mass going to state $x_{1,i}$. Each state transforms at a random rate $k_{1,i}$. Mass incoming to pool 2 is partitioned into states $\mathbf{x}_2$ with mass fractions specified by $\mathbf{p}_2$. All states decay at the random rates $k_{2,i}$ to CO_2. The system matrix $\mathbf{A}$ for distributed serial processes is lower triangular.

first-order differential equations which includes all states

$$\begin{aligned} \frac{\mathrm{d}x_{1j}}{\mathrm{d}t} &= -k_{1j}x_{1j} + J(t)p_{1j}, \quad j = 1, \ldots, n \\ \frac{\mathrm{d}x_{2j}}{\mathrm{d}t} &= -k_{2j}x_{2j} + p_{2j}\sum_{i=1}^{n} f_{1i}k_{1i}x_{1i}, \quad j = 1, \ldots, n. \end{aligned} \tag{4.11}$$

Here, x_{1j} is the mass associated with the jth state in carbon pool 1, and x_{2j} is the mass associated with the jth state in carbon pool 2. The probability p_{ij} partitions mass incoming to the ith pool into the state j and has values such that $\sum_j p_{ij} = 1$. The conversion fraction, f_{ij}, is associated with microbial transformation, representing the probability that carbon involved in the j^{th} process will be transformed to a microbial component, while $(1 - f_{ij})$ is the probability that the carbon will be respired to CO_2. The conversion fraction f can also be associated with more general losses from the system, such as the leeching of carbon to groundwater from certain states, or abiotic oxidation such as photo-oxidation or ozone-oxidation [140].

We can write this dynamics of the linear system (4.11) in matrix form, just as we did in equation (4.4), as

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x} + J(t)\mathbf{p}_0, \tag{4.12}$$

where $J(t)$ is the input of fresh dead organic matter to the system and $\mathbf{x}$ and $\mathbf{p}_0$ are the concatenated vectors

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \text{ and } \mathbf{p}_0 = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{0} \end{bmatrix}, \mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_{11} \\ \vdots \\ \mathbf{x}_{1j} \\ \vdots \\ \mathbf{x}_{1n} \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} \mathbf{x}_{21} \\ \vdots \\ \mathbf{x}_{2j} \\ \vdots \\ \mathbf{x}_{2n} \end{bmatrix}, \mathbf{p}_1 = \begin{bmatrix} \mathbf{p}_{11} \\ \vdots \\ \mathbf{p}_{1j} \\ \vdots \\ \mathbf{p}_{1n} \end{bmatrix} \tag{4.13}$$

The components of $\mathbf{x}_i$ of pool i are the states $x_{i,j}$ and the components of $\mathbf{p}_1$ are p_{1j}. The vector $\mathbf{p}_0$ represents the initial fractionation of incoming fresh dead organic matter into the network. The vectors $\mathbf{x}$ and $\mathbf{p}_0$ have length z, the total number of states in the system.

Equation (4.12) describes more generally any network of y pools connected with distributed rates. In the more general case, $\mathbf{x} = [\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_y]^T$ represents the states within all pools and $\mathbf{p}_0$ contains the partition vectors $\mathbf{p}_i$ of all pools which initially receive mass, rescaled so that $\sum_j \mathbf{p}_{0,j} = 1$

The solution to this system when the input $J(t)$ is an initial impulse of size G_0 is

$$\mathbf{x} = G_0 \mathbf{U} e^{-\mathbf{\Lambda} t} \mathbf{U}{-1}\mathbf{p}_0 \tag{4.14}$$

as derived in Appendix 4.10.2. The matrix $\mathbf{U}$ contains the eigenvectors $\mathbf{u}_i$

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_z, \end{bmatrix}, \tag{4.15}$$

$\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_z. \end{bmatrix}, \tag{4.16}$$

and $e^{-\mathbf{\Lambda} t}$ is the matrix exponential. The total mass remaining in the decay system is

$$G(t) = \sum_i x_i(t). \tag{4.17}$$

Substituting (4.14) for $\mathbf{x}$ in equation (4.17) provides the dynamics of the total mass remaining in the system as a function of time

$$g(t) = \sum_i r_i e^{-\lambda_i t}, \tag{4.18}$$

where $g(t)$ is the mass fraction remaining in the system $G(t)/G_0$. r_i are the elements of the vector

$$\mathbf{r} = \mathbf{S}\mathbf{U}^{-1}\mathbf{p}_0 \tag{4.19}$$

which partitions the incoming mass into the parallel decay states. Because the mass in the system is the sum of all states, the matrix $\mathbf{S}$ represents the sum of each eigenvector $\mathbf{u}_i$ (see appendix 4.10.3)

$$\mathbf{S} = \begin{bmatrix} \sum_j u_{1,j} & 0 & \dots & 0 \\ 0 & \sum_j u_{2,j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_j u_{z,j} \end{bmatrix}. \tag{4.20}$$

Because $\mathbf{r}$ derives from the projection of the initial state vector $\mathbf{p}_0$ onto the eigenvectors, there is no reason for the elements of the projected vector $\mathbf{r}$ to remain positive. We find that these negative components are related to a time lag rather than a decay. Furthermore, there is no constraint on the system $\mathbf{A}$ to have real eigenvalues. However because eigenvectors come in complex conjugate pairs, values of r_i cannot be complex. Decomposition systems however do appear to be dissipative enough that complex eigenvalues are not prevalent; we return to complex eigenvalues in our discussion of feedback in section 4.4.3.

The stochastic description of carbon dynamics aids the interpretation of carbon movement throughout the decomposition network and the Markov representation of the system has some important properties equivalent to those calculated from the deterministic solution. The probability P_{ij} that a particle in state i at time 0 will be in state j at time t is [88, 141]

$$\mathbf{P}(t) = e^{\mathbf{A}t} = \mathbf{U}e^{\mathbf{\Lambda}t}\mathbf{U}^{-1} \tag{4.21}$$

where $\mathbf{P}$ is the $z \times z$ matrix with elements P_{ij}, $\mathbf{A}$ is the same matrix in equation (4.12) that contains the transition or transformation rates k. $\mathbf{U}$ and $\mathbf{\Lambda}$ contain the eigenvectors and eigenvalues of $\mathbf{A}$.

Similarly, the survival probability at time t of particles which were initially distributed amongst the states $\mathbf{p}_0$ at $t = 0$ is equivalent to the impulse response $g(t)$ given by equation (4.18) [90, 88]. An additional property is that the mean residence time of particles in state j which originated in state i is given by the elements of the matrix A^{-1} [88, 141].

4.3 Obtaining the Exit Rate Function $r(k)$ from a Network

If we assume that transformations between pools occur at a continuous distribution of rates, it is natural to expect there to be a continuum of exit rates $r(k)$ from the system. The mass remaining in the system is then described by the continuous superposition

$$g(t) = \int_0^\infty r(k)e^{-kt}dk, \tag{4.22}$$

which is the continuous version of equation (4.18). Details of the extrapolation of the network model to the continuum can be found in appendix 4.10.5.

Equation (4.22) reveals that the mass lost from a decay system can be interpreted as the Laplace transform of the initial diagonalized state function, or exit rate function $r(k)$. It indicates that the parallel description of organic matter decay is simply a projection of the complete system onto the coordinates $\mathbf{U}$. In this manner, the impulse response is described in terms of the eigenvalues, or poles [142, 143], of the underlying decay system. Equation (4.22) is equivalent to the original parallel decay equation (4.1) except that $r(k)$, like $\mathbf{r}$, may be negative. Ultimately, we seek to invert decay data in order to find the exit rate distribution $r(k)$ for a litter decay data set. But first, we calculate and interpret $r(k)$ for various examples of network configurations.

4.4 Example Systems

In this section we calculate and interpret $r(k)$ for four examples of network configurations: 1) serial networks, 2) feedback networks, 3) the CENTURY degradation model which contains parallel, serial and feedback processes, 4) and an amorphous network containing a continuum of interconnected states. These examples help to develop an intuition for interpreting $r(k)$.

4.4.1 Calculating $r(k)$ from a network with heterogeneous states and distributed rates

In each of these cases, we prescribe a network configuration and calculate the exit rate distribution $r(k)$ associated with it. To numerically determine $r(k)$ associated with a given network, we use a Monte-Carlo simulation with N trials in order to estimate a mean $r(k)$ as discussed in Appendix 4.10.6. Because we assume rate heterogeneity is wide, we change variables from $k \rightarrow \ln k$ and calculate $v(\ln k)$, which is equivalent to $r(k)$ in $\ln k$ space. $v(\ln k)$ and $r(k)$ are related by the equation $v(\ln k) = r(k)\frac{\mathrm{d}\ln k}{\mathrm{d}k}^{-1}$. The mass fraction remaining expressed in terms of $v(\ln k)$ is

$$g(t) = \int_{-\infty}^\infty v(\ln k)e^{-kt}d\ln k. \tag{4.23}$$

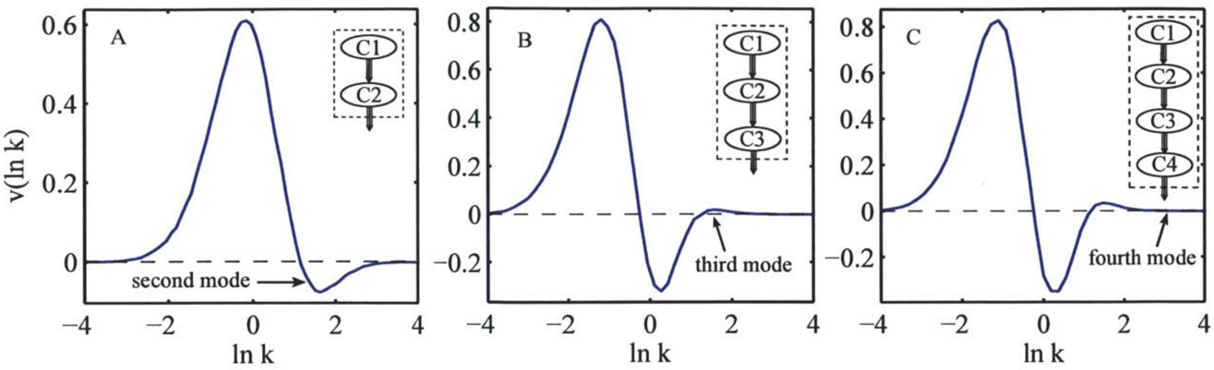

Figure 4-3: Initial eigenstate distributions $v(\ln k)$ for distributed serial processes. To generate each plot, each pool is fractionated amongst 32 states that decay at rates randomly chosen from a distribution $p_i(k)$ associated with the i^{th} pool. We estimate the average initial states function $v(\ln k)$ using the Monte-Carlo procedure given in section 4.3. All mass is allocated in the first pool at $t = 0$ and mass only exits the system from the final pool in the series; $f = 1$ for all other pools. A: The initial distribution $r(k)$ corresponding to two pools in series. Carbon leaves the first pool at a distribution of rates $p_1(k)$ and leaves the second pool according to the distribution $p_2(k)$. $p_1(k)$ is lognormal with a mean and variance of $\log k$ equal to 0 and 1 respectively. $p_2(k)$ is lognormal with a mean and variance of $\log k$ equal to -1 and 1 respectively. The rescaled sum $v(\ln k)$ of 8000 realizations of $\mathbf{r}$ are plotted here. B: Three pools in series. Flows out of each pools are set by the distributions $p_1(k)$, $p_2(k)$, and $p_3(k)$. The mean and variance of $\log k$ for each distribution is as follows: p_1 : 0 and 1, p_2 : -1 and 1, p_3 : 1 and 1. The sum of 10000 instances of $\mathbf{r}$ are plotted here. C: Four pools in series. Flows out of each pool are given by the same distributions p_1, p_2, p_3 in Figure B, with the flow out of the fourth pool set by p_4, which is lognormal with a mean and variance of $\log k$ equal to 3 and 1. The sum of 10000 instances of $\mathbf{r}$ are plotted here.

4.4.2 Serial pools

Here we calculate and interpret $r(k)$ resulting from multiple-pools connected in series. We first consider two pools in series as shown in Figure 4-2. The mass entering each pool is partitioned into different rates by the distributions $p_1(k)$ and $p_2(k)$. In this and future examples we choose the distributions to be lognormal. We assume this based on the finding that decay rates are lognormally distributed [13] and the suggestion that organic matter quality and transformations may be lognormally distributed [144]. We then use a Monte-Carlo simulation to randomly sample 32 rates from each distribution, assemble the connectivity matrix $\mathbf{A}$, and then calculate its eigenvalue decomposition and associated $\mathbf{r}$. We repeat this process 10000 times to generate an ensemble of initial eigen-decay state vectors $\mathbf{r}$. We follow the procedure outlined in appendix 4.10.6 to estimate the ensemble average $v(\ln k)$. Figure 4A reveals that the average $v(\ln k)$ is smooth, has one inflection point, and becomes negative in the faster portion of the spectrum. The positive region of $v(\ln k)$ resembles the lognormal distribution associated with the slower transformations in the serial network.

For any set of processes purely in series, as shown in Figure 3B, the full system matrix $\mathbf{A}$

can be written as a lower triangular matrix, because information only flows in one direction. A relevant property of the triangular matrix $\mathbf{A}$ is that its diagonals, which represent the rates of state transformations, are also its eigenvalues. Therefore the exit rates are closely related to the underlying transformation rates in the system.

Because mass must pass through both a slow and fast pool before exiting, a negative mode appears in $r(k)$ around the rates of the fast pool. This occurs because the system eigenvalues are still the diagonals of $\mathbf{A}$, but matter cannot exit the system at the rapid rates of the fast pool. To compensate, the weights $r(k)$ take on negative values at these faster rates, indicating a lag or delay at those time scales.

We repeat this process for systems containing three and four lognormal pools in series with varying mean and variance of $\ln k$. We find that the number of modes can reflect the number of kinetically distinct compartments in the network. The sign of the modes however is related to the mass loss, or conversion fraction f, associated with transfers from that pool. If enough mass exits the system from a pool, the function $v(\ln k)$ may remain positive over the rates $\ln k$ associated with that pool. If little mass is lost and more mass is transferred at rates $\ln k$, negative modes may be associated with those rates. The negative mode signifies a lag time associated with those processes. Because rate distributions of pools may overlap, the number of modes in the distribution $r(k)$ indicates a lower bound on the number of macroscopic processes in series. The exact number of serial processes can be estimated by using the "phase function method" [132] when decay data has high resolution near $t = 0$ [145].

This approach fails however, as do the standard popular compartmental terrestrial decomposition models, when there is a long lag time, or inoculation time, before the onset of decomposition. A long lag time may result from highly serial processes, or when no transformations occur while while an organic substrate is being inoculated. Purely serial processes with random rates have a generalized Erlang distribution of residence times [90, 91]. Because of the large lag or delay time associated with long chains of purely serial processes, Monte-Carlo simulations show large rapid fluctuations in the weights of the projected eigen-decay states $\mathbf{r}$ and fail to generate a stable mean $r(k)$. However, if a large lag time actually represents an inoculation time where no transformations occur, modeling this waiting time via many serial transformation processes is likely inappropriate. Rather, we suggest modeling the inoculation time with a simple lag time parameter, a, by transforming time to $t^* = t - a$.

4.4.3 Feedback processes

Feedback loops are fundamental to decomposition systems since decomposers recycle old organic matter into new biomass that remains part of the soil organic matter decomposition system. We suggest that feedback between carbon pools takes place stochastically at different rates, as shown in Figure 4-4. In this section we consider simple arrangements of pools in feedback to generate the

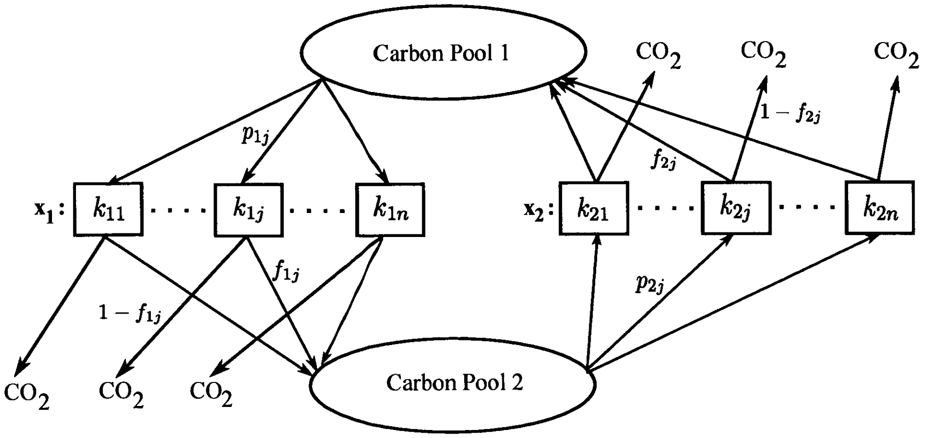

Figure 4-4: A feedback process between two pools with distributed rates. Organic matter may be mineralized or otherwise lost during some processes. Ovals represent macroscopic pools of organic matter, while boxes represent the fractionation **p** of the preceding pool into different states **x** which decay at different rates k. Mass fractions f are transferred to the next pools, while the fractions $1 - f$ are lost from the system.

system matrix **A** and calculate its eigenvalue decomposition, using the same procedure outlined in section 4.3, to identify the exit rates associated with feedback decay.

The exit rate functions of three different feedback loops are shown in Figure 4-5. We first consider two pools in series connected by two lognormal distributions of rates p_1 and p_2, with a conversion fraction $f = .5$ associated with each transformation. The rates drawn from p_1 are two natural orders of magnitude (a factor of e^2) slower than the rates drawn from p_2. In the first case, Figure 4-5A, incoming mass first enters the slowly decaying pool and then feeds into the faster pool. The exit rate function $v(\ln k)$ is shifted slightly left of p_1(which has a mean $\ln k$ of 0) since particles may cycle through the system more than once before exit. In Figure 4-5B, the pools are switched with the fast pool decaying first followed by the slow pool. Because 50% of the mass exits after the fast pool, the signal from the fast pool becomes strong, appearing as a second mode in the exit rate function. The order of transformations may therefore influence the eigenstate distribution $r(k)$. Figure 4-5C shows the exit rates associated with three pools in series, with 70% of mass retained at each transformation (30% lost to CO2). The medium and slow modes dominate, while the fast mode highlights the serialilty of the system.

Systems with feedback have a full **A** matrix and eigenvalues which may be complex [18]. Decays with complex eigenvalues take the shape of an exponentially damped oscillation. Real components of the eigenvalues represent exponential decay rates, while the imaginary components represent the frequencies of sinusoidal oscillations. However, in the test models of Figure 4-5, the system matrices **A** are diagonally dominant. Most eigenvalues approximate the diagonals of **A** and less than 1% are actually complex. Furthermore, those which are complex typically have an imaginary

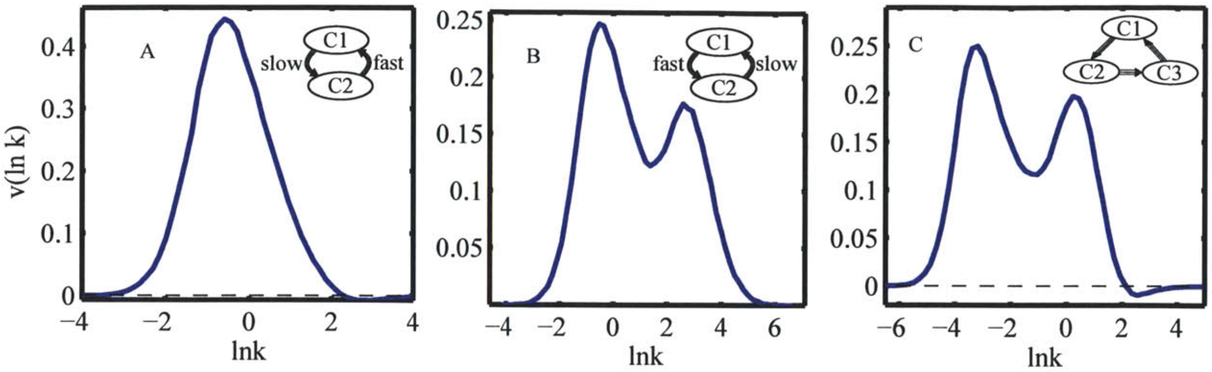

Figure 4-5: Exit rate functions associated with feedback processes. A: Two carbon pools in feedback. Pool C1 decays with 32 random rates sampled from a distribution p_1 and pool C2 decays with 32 random rates sampled from a distribution p_2. p_1 and p_2 are both lognormal with a mean and variance of log rates as follows, p_1: 0, 1 ; p_2 : 2.5, 1. All states have the same conversion fraction $f = 0.5$. Only a few eigenvalues are complex, indicating oscillations in some eigenstates, however in those cases, imaginary components were less than 5% of the decay rates, as discussed in the text. This figure uses an ensemble of 8000 randomly generated trials of $\mathbf{r}$. B: Two pools in feedback, with rate distributions p_1 and p_2 opposite of Figure A. The mean and variance of log rates of these distributions are p_1: 2.5, 1; and p_2: 0, 1. All states have conversion fraction $f = 0.5$. Changing the order of C1 and C2 strongly influences $v(\log k)$. 4000 ensembles were used. C: Three carbon pools in feedback. The three distributions p_1, p_2, p_3 are lognormal with a mean and variance of $\log k$ as follows, p_1 : 0,1 ; p_2 : 2.5, 1; p_3: −2.5, 1. All states have conversion fraction $f = 0.7$. 4000 ensembles were used.

component, $\Im(\lambda)$, less than 10% of the real component $\Re(\lambda)$, indicating a damping ratio [142]

$$\zeta = \sqrt{\frac{1}{\left(\frac{\Im(\lambda)}{\Re(\lambda)}\right)^2 + 1}} \approx 1$$

These decays are approximately critically damped [142] and exhibit essentially no oscillation. Including the oscillatory components in the solution only changes the mass loss by a negligible amount and therefore only the real components of the eigenvalues are plotted on the horizontal axis of Figure 4-5. A multivariate pdf of $r(k)$ in complex rate space could be plotted; however the imaginary components do not reveal any further information about the decay in such systems and only complicate the picture.

Real soil carbon systems are also highly dissipative [7] as the conversion fraction, or substrate utilization (carbon uptake) efficiency, is small to moderate such that $0.1 < f < 0.7$ [33, 41, 7, 8], indicating that significant mass losses are associated with transformations. This property is reflected in calibrated models of decomposition networks [46, 93, 44]. As a result, both the Roth C and CENTURY models exhibit real eigenvalues for a wide range of system parameters [18]. We now proceed to investigate the CENTURY model with heterogeneous compartments.

4.5 Assembling Compartmental Models with Distributed Rates

Larger compartmental models with heterogeneous kinetics can be created by connecting pools to each other in parallel, series, and feedback with a distribution of rates. Contemporary compartmental models of soil organic matter decomposition [45,44,93] typically connect pools with a single rate constant and do not account for a continuum of transformations.

By fractionating each compartment into a distribution of particle-states, the residence time of carbon in the pool is no longer exponentially distributed [139,89]. In fact, a compartment with any residence time distribution can be approximated by fractionating that compartment into an appropriate arrangement of sub-states in series, parallel and and/or feedback [88,90]. The residence time distribution associated with the passage through a network of linear compartments, Markov process, etc. is called a distribution of phase type [88,90,91]. Compartmental models containing pools with arbitrary residence time distributions are called semi-Markov models [88,89]. While we've assumed in this and prior sections that each compartment contains particle-states which are in parallel, the inversion method presented in section 4.7, allows for more general particle-state configurations.

4.5.1 Exit rates from the CENTURY model

We now investigate the exit rate function of the CENTURY model [45], assuming each compartment transfers carbon with a distribution of rates. In this modification of CENTURY, transfers occur between five carbon pools. Likely ranges of turnover times are suggested for each pool [45]: (1) structural litter [1-5yr], (2) metabolic litter [.1 - 1yr], (3) active soil [1-5yr], (4) slow soil [20-40yr], (5) passive soil [200-1500yr]. Their inverse provides a suggested range of turnover rates. Although pools likely have dispersed kinetics, the CENTURY model is highly parametrized in order to specify exact single rates of transfer between pools as a function of soil, litter type, and environmental conditions. Previously, an eigenvalue decomposition of the CENTURY model has revealed the model to behave as a parallel system [18] for most conditions, seldom having negative values of $\mathbf{r}$ and complex rates λ.

Figure 4-6 shows exit rate functions from the CENTURY model [45] assuming pools have heterogeneous kinetics. The exit rate function associated with case 1 is calculated by randomly assigning lognormal rates to 32 sub-compartments in each pool, assuming that the range of suggested turnover rates represent four standard deviations of $\log k$. The mean of $\log k$ for each pool is assumed at the center of the suggested ranges. We use a Monte Carlo simulation as described in appendix 4.10.6 to calculate the exit rate function $r(k)$ from $N = 5000$ trials. The exit rate function $v(\ln k)$ given these assumptions, case 1, is shown as the green line in Figure 4-6. There are five pools, but only four modes are present in $v(\ln k)$ because pools 1 and 3 have the same suggested turnover times and therefore have similar eigenvalues. We then increase the variance of rates within the compartments and re-calculate the exit distribution to show what happens as heterogeneity increases within each

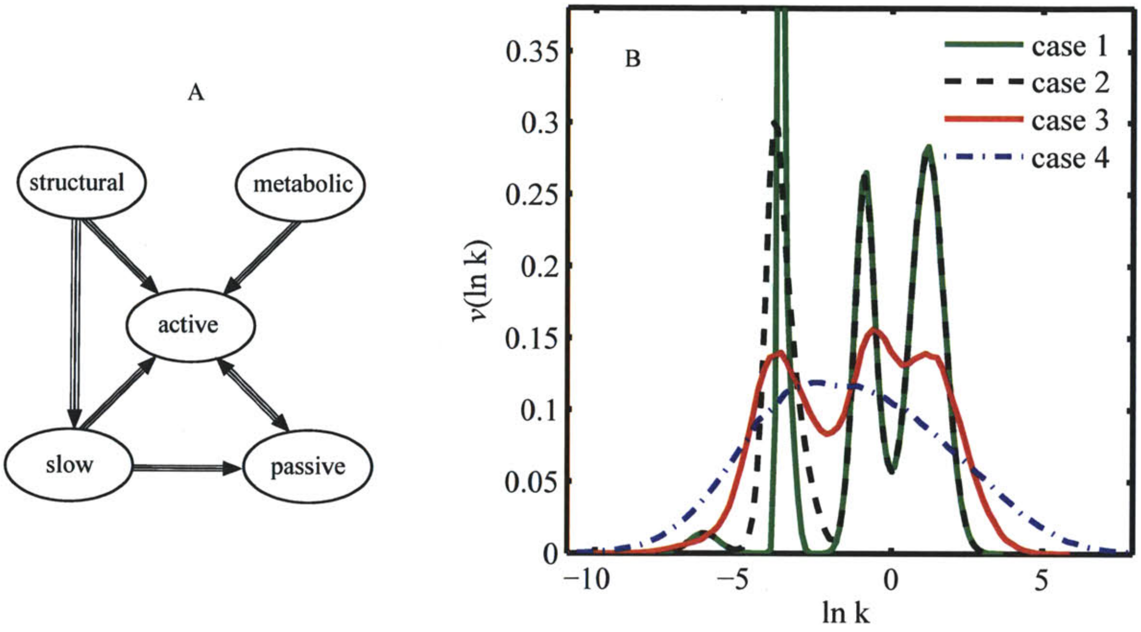

Figure 4-6: Exit rate function from the CENTURY [45] model assuming pools have heterogeneous kinetics. Figure A shows the architecture of the CENTURY Model. Figure B shows $v(\ln k)$ associated with four different choices of the rate variances σ_i. We calculate $v(\ln k)$ using the Monte-Carlo method described in section 4.3. For each case, we assume each pool has rates which are lognormally distributed and set the mean log rate of each pool μ_i equal to the mean of the log of the turnover rates suggested in the text. For all cases, $\mu_1 = -0.80, \mu_2 = 1.15, \mu_3 = 0.80, \mu_4 = -3.34, \mu_5 = -6.31$. The variance of each pool is set as follows. Case 1, solid green line: the standard deviation for each pool σ_i is set equal to the range of log rates for each pool divided by four: $\sigma_1 = 0.40, \sigma_2 = 0.58, \sigma_3 = 0.40, \sigma_4 = 0.17, \sigma_5 = 0.50$. The range of suggested rates therefore span four standard deviations. Case 2, dotted black line: σ_i are the same as case 1 except σ_4 is increased to $\sigma_4 = 0.52$ roughly the same size as the other pools. Case 3, dashed red line: all σ_i are increased by a factor of two, $\sigma_1 = 0.80, \sigma_2 = 1.15, \sigma_3 = 0.80, \sigma_4 = 1.04, \sigma_5 = 1.00$. Case 4, dot dashed blue line: σ_i are further increased: $\sigma_1 = 1.61$, $\sigma_2 = 2.30$, $\sigma_3 = 1.60$, $\sigma_4 = 2.08$, $\sigma_5 = 2.01$. The range of suggested rates spans one standard deviation. For all cases, we assume litter has a lignin fraction of 0.18, nitrogen fraction of 0.05, a soil with a combined silt and clay content of 0.75, and a temperature and moisture factor of 1. These values set the other parameters in the CENTURY model [45].

pool. The resulting exit rate distributions associated with three changes in the variance are shown in Figure 4-6. As the variance within each compartment increases, the properties of specific individual transformations in the network become less important and for wide heterogeneity, only a smooth continuum of rates prevails.

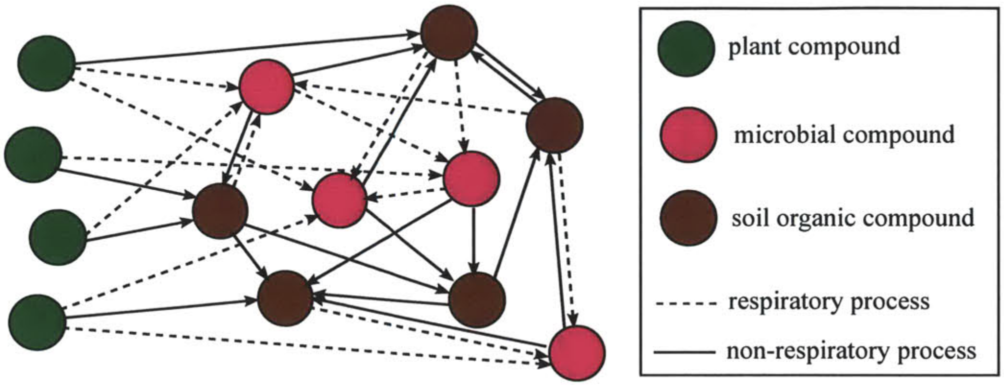

Figure 4-7: Transformation network between randomly configured litter, microbial, and soil organic matter states. Matter is initially allocated in litter states and is transformed to either microbial or other soil organic matter states. Transformation of carbon compounds to microbial states requires energy, and a fraction of carbon is converted to microbial structures via biosynthesis pathways, while the remaining carbon is oxidized to carbon dioxide or other inorganic products via respiration pathways. While transformations between soil organic matter states may be microbially mediated, this carbon is not processed internally by the microbial cell and therefore does not take part in respiration.

4.6 Amorphous Networks

In this section, we remove the concept of pools and consider soil as a continuum of plant matter, microbial, and other transformed soil organic matter states. This is depicted in Figure 4-7. Initial litter states may transform to microbial or soil organic matter states at rates given by the distribution ρ_l, microbial states transfer to other microbial and soil organic matter states at rates given by ρ_m, and soil organic matter states transfer to other soil organic matter states and microbial states at rates given by ρ_s.

We calculate the exit rate function $v(\log k)$ for a network composed of a number of litter, microbial, and soil organic matter states states. Processes leading to a microbial state have a conversion fraction $f = 0.4$, consistent with estimates of substrate utilization (carbon uptake) efficiency [33, 41, 7, 8]. The rates associated with litter, microbial, and soil organic matter states are randomly pulled from the lognormal distributions ρ_l, ρ_m, and ρ_s with parameters specified in the caption of Figure 4-8. States are also connected randomly for each trial to generate the system matrix $\mathbf{A}$. As described in section 4.2.2, we perform an eigenvalue decomposition to find the exit rate vector $\mathbf{r}$ for $N = 10000$ random configurations of rates and connections in the network. We plot the rescaled sum of all exit rates $v(\log k)$ in Figure 4-8. Again we find that the $\mathbf{A}$ matrix is diagonally dominant, leading to eigenvalues which are approximately the same as the rates of the state transformations themselves. For this particular choice of parameters $v(\log k)$ is roughly unimodal.

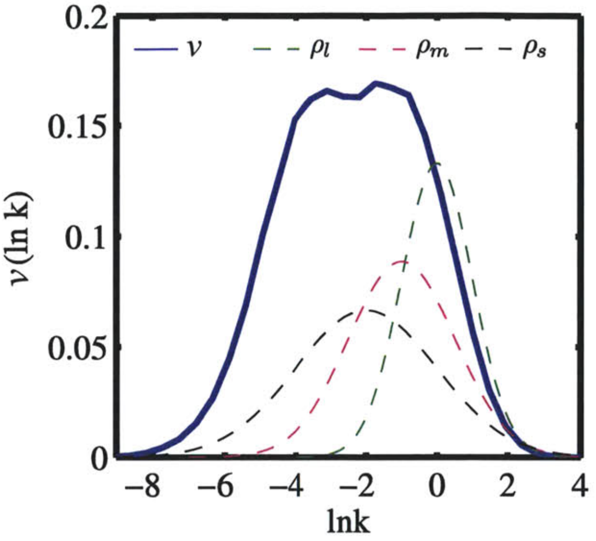

Figure 4-8: Exit rate function $v(\ln k)$ for $N = 10000$ randomly configured networks consisting of 32 litter states, 32 microbial states, and 32 soil organic matter states. Mass flows initially from litter to microbial and soil organic matter states. Any flow to a microbial state has a conversion fraction $f = 0.4$ of conversion to microbial biomass. Litter states are randomly assigned from a lognormal distribution ρ_l (green dashed line) having a mean and standard deviation of log rates of 0 and 1, respectively. Microbial states are randomly assigned from a random distribution ρ_m (magenta dashed line) having a mean and standard deviation of log rates of -1 and 1.5, respectively, and soil organic matter states are assigned from a random lognormal distribution ρ_s (black dashed line) having a mean and standard deviation of log rates of -2 and 2, respectively. The distributions ρ_l, ρ_m, and ρ_s have been rescaled in the y-direction for clarity.

4.7 Analyzing LIDET Data

The previous sections have discussed how the function $r(k)$ relates to an underlying heterogeneous linear decay network $\mathbf{A}$. The exit rate function $r(k)$ contains all of the information regarding mass loss in such a network. Unfortunately, actual degradation networks are complex, and modeling them well enough to estimate a complete picture of $r(k)$ is nearly impossible. With only limited knowledge of the mechanisms and network structure, attempts at characterizing these networks require a large amount of empirical studies, approximations, heuristics, intuition, and curve fitting [46,45]. Instead we suggest simply identifying $v(\ln k)$. By doing so, we characterize only the poles or eigenvalues of the decay network.

We obtain $r(k)$ by inverting mass loss data from the LIDET dataset. The inversion is constrained so that $r(k)$ can represent a network of serial, parallel, and feedback processes. We proceed with the log-transformed version of the exit rate function $v(\ln k)$ and equation (4.23). Because equations (4.22) and (4.23) are a Laplace transform, the function $v(\ln k)$ is obtained by calculating the inverse Laplace transform of $g(t)$. However, the inverse Laplace transform is ill-posed [124, 115, 112], meaning solutions $v(\ln k)$ are highly sensitive to data noise [51]. Regularization methods

[124, 115, 112] are commonly used to calculate the inverse Laplace transform of noisy data by seeking solutions with minimal degrees of freedom. Here we use Tikhonov regularization [115, 116] to identify an optimally smooth solution that best fits the data. We have previously applied this technique to equation (4.1) in order to identify the rate probability distributions $p(k)$ associated with litter decay data sets [13, 51]. Here, $r(k)$ is not required to be a probability distribution, but its shape is nevertheless physically constrained. The physical constraints on the decay system are that the system mass cannot be negative,

$$g(t) \geq 0 \tag{4.24}$$

and the mass cannot increase

$$dg(t)/dt \leq 0. \tag{4.25}$$

To efficiently implement these constraints, we use the approximations

$$g(t) > 0 : \int_{-\infty}^{K} v(\ln k) d \ln k > 0 \tag{4.26}$$

$$dg(t)/dt < 0 : \int_{-\infty}^{K} kv(\ln k) d \ln k < 0. \tag{4.27}$$

When $v(\ln k)$ is not very wide, the constraints (4.26), (4.27) are conservative approximations to (4.24) and (4.25). A proof is provided in Appendix 4.10.7. When $v(\ln k)$ is wide, e^{-kt} in (4.23) approximates a low-pass filter on $v(\ln k)$ at the rate $K = 1/t$; setting the integration limit to $K = 1/t$ approximates $v(\ln k)e^{-kt}$ well when $v(\ln k)$ is wide. With these approximations in mind, we evaluate the integral (4.27) at many limits $K = K_i$ within the bounds of the function $v(\ln k)$ (see Appendix 4.10.7) to ensure mass does not increase at any time t. Although the constraints (4.24) and (4.25) could be exactly evaluated, we implement constraints (4.26) and (4.27) for computational expedience.

4.7.1 Inversion of a LIDET dataset

An inversion $v(\ln k)$ of a LIDET dataset is shown as the red line in Figure 4-9A. There appears to be one inflection point, becoming negative at fast k. In the negative region, mass is not exiting the system at those rates; rather the system exhibits lag or serial transfer at those rates because more mass is sequentially transforming at those rates than exiting the system.

The exit rate function $v(\ln k)$ in Figure 4-9A may derive from a dissipative network comprised of a continuum of states which transform and decay at a continuum of rates. It may also similarly derive from a network of heterogeneous compartments. We test the hypothesis that this exit rate function results from by two or three heterogeneous pools in series. We first consider a two pool serial network with lognormally distributed rates. The rates of transformation from pool 1 to pool

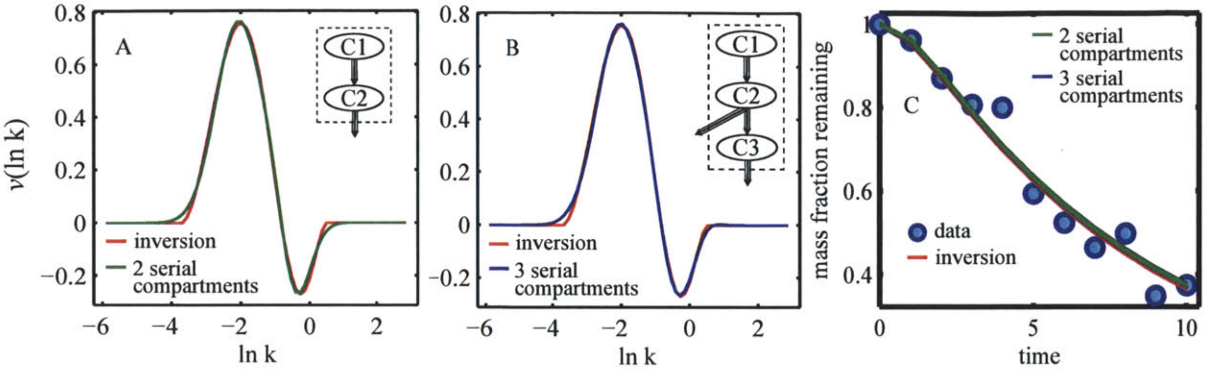

Figure 4-9: Exit rate function $v(\ln k)$ associated with a LIDET dataset. A: Red line is the exit rate function calculated from the LIDET dataset in C. Green line is the exit rate function corresponding to the best fitting 2 compartment serial network with lognormally distributed transformation rates. B: Red line is the same in A. Blue line is the exit rate function from the best fitting 3 compartment serial network with lognormally distributed transformation rates. C: LIDET mass loss data (circles), mass loss prediction from the inversion (red), 2 serial compartments (green), 3 serial compartments (blue).

2 are described by the lognormal parameters μ_1 and σ_1. Mass leaves the system from pool 2 at lognormally distributed rates parametrized by μ_2 and σ_2. We assume mass transfers between pool 1 and pool 2 with a single conversion fraction f, although a distribution of conversion fractions f_{1j2} associated with each state could also be assumed. We then use Matlab's 'nlinfit.m' function to fit the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2$, and f to the to the exit rate function $v(\ln k)$ estimated from the decay data. We find the parameters $\mu_1 = -0.18$, $\sigma_1 = 0.42$, $\mu_2 = -2.11$, $\sigma_2 = 0.63$, $f = 1.0$ best fit the $v(\ln k)$ from the inversion. The exit rate function corresponding to these parameters is plotted as the blue line in Figure 4-9A. The conversion fraction $f = 1$ physically suggests the first compartment represents some distribution of inoculation rates, since all mass transfers to a second decomposition state without loss. Decay then proceeds heterogeneously in the second compartment. Because this transformation is loss-less, the pools may be reversed; the inoculation processes may be slow, whereas the inoculated decomposing state may decay fast. Since these compartments have heterogeneous kinetics, each compartment itself may represent a complex sub-network of reactions, transformations, and or other processes [88].

We then consider a three pool serial model with lognormal transformations. Mass exits the first pool at lognormal rates with parameters μ_1 and σ_1 and may transform to the second pool with conversion fraction f_{12} or to the third pool with conversion fraction f_{13}. Mass exits the second pool at lognormal rates with parameters μ_2 and σ_2 and transforms to the third pool with conversion fraction f_{23}[1]. Mass exits the third pool at lognormal rates with parameters μ_3 and σ_3. The parameters which best fit the inversion $v(\ln k)$ are $\mu_1 = -0.12$, $\sigma_1 = 0.50$, $\mu_2 = -2.11$, $\sigma_2 =$

---

[1] Again, we assume here for simplicity that the conversion fractions f_{ijl} of all states x_j associated with transfers from pool i to pool l are constant and use the shorter notation f_{il}.

0.64, $\mu_3 = 0.56$, $\sigma_3 = 0.40, f_{12} = 0.99, f_{13} = 0, f_{23} = 0.20$. Pools 1 and 2 from the three pool network have similar properties as pools 1 and 2 from the two pool network and represent a distributed inoculation time pool followed by a heterogeneous decomposition pool. No mass is transferred from pool 1 to pool 3. However, the parameter fitting procedure identifies that 20% of the mass exiting the second pool enters a third pool having turnover times distributed between approximately 90 and 450 days. These turnover times are consistent with measured estimates of soil microbial turnover having substrate utilization efficiency of about 0.2 [146]. We therefore suggest this pool is microbial. The exit rate function calculated from the three pool network is the blue line in Figure 4-9B. It provides a marginally better fit to the inversion than the two pool network of Figure 4-9A. Even more realistic values may be encountered if allow for feedback between pool 2 and 3.

The best fitting distributions of transformation rates between the two and three pools could be calculated using an additional inversion technique. However, this approach may be excessive as 1) the true underlying network architecture remains unclear and 2) the exit rate function $v(\ln k)$ itself contains all information regarding mass loss from the network. Nevertheless, understanding the distributions of the underlying transformations might provide additional clues to how environmental and compositional drivers affect the overall exit rate function $v(\ln k)$.

4.7.2 Inversion of 232 LIDET datasets

We proceed to calculate the regularized inverse for 232 LIDET datasets. We consider only the 237 LIDET datasets which contain 5 data points with replicates. Of those 237 we find 5 have insignificant mass loss after the first data point and we do not attempt to describe those datasets [13]. We find that the 232 inversions are either unimodal, bimodal, trimodal, or quadmodal. Examples of each are shown in Figure 4.7.2. We find that 38 are unimodal, 170 are bimodal, 12 are trimodal, and 7 are quadmodal. Of the 38 unimodal solutions, 14 contain just one rate and decay exponentially. These are determined by calculating the non-negative inverse of the data (unregularized) [51, 13]. We then check whether that inverse contains just one rate and fits the data better than the regularized estimate of $v(\ln k)$.

The 12 trimodal $v(\ln k)$ can all be approximated by a simpler solution. Because the third mode of these datasets is a positive mode located at extremely fast rates, the entire mode decays to a negligibly small fraction by the time of the first measurement. In fact, for all of these solutions, the mass associated with this mode has decreased to approximately the double precision limit by the first data point and is effectively zero [51]. Therefore, this mass is experimentally equivalent to an instantaneous loss or leaching of mass. We thus suggest that the mass of the third mode is most simply represented by an instantaneous leaching of mass at $t = 0$ [51]. The inversion and its approximation (with the mass located at an infinite rate constant) is shown in Figure 4-11A. Figure 4-11C shows that the decay predicted from the trimodal $v(\ln k)$ and the leached-fraction

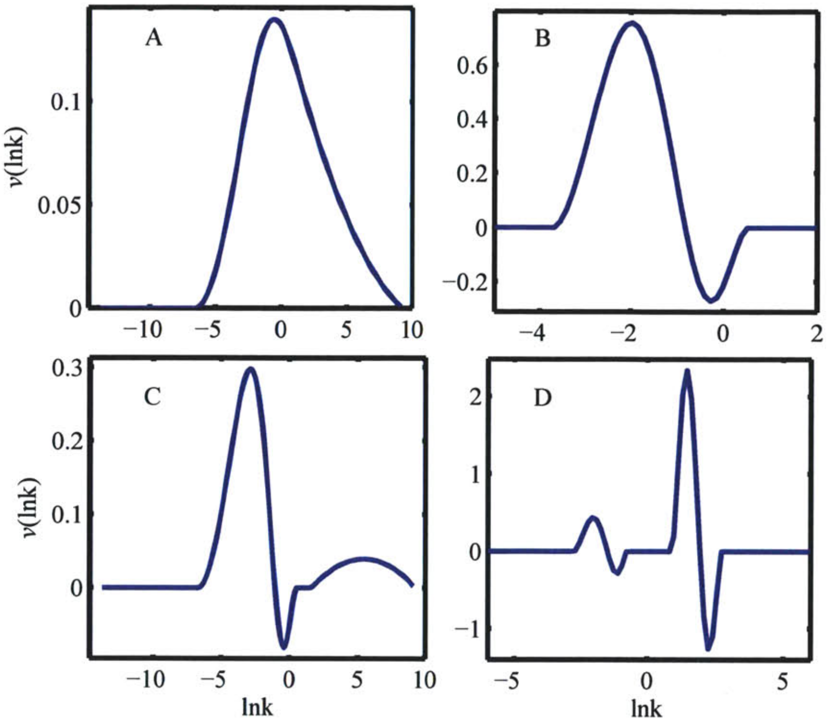

Figure 4-10: Four different classes of exit rate functions $v(\ln k)$ found amongst 232 LIDET datasets. A: Unimodal (38 datasets). B: Bimodal (170 datasets). C: Trimodal (12 datasets). D: Quadmodal (7 datasets).

approximation provide an equivalent prediction of mass loss over the duration of the experiment.

Most of the quadmodal results can also be simplified. Six of the seven datasets contain mass in two slow modes that exhibit negligible decay over the duration of the experiment. The total mass of these slow modes can therefore be equivalently represented by a constant inert mass fraction. An example of a quadmodal distribution and its inert mass fraction approximation (located at the slowest rate in the domain) is shown in Figure 4-11B. Figure 4-11D shows the quadmodal $v(\ln k)$ and its inert-fraction approximation predict the same decay over the duration of the experiment.

In our previous study [13] we identified the qualitative effects of climatic and compositional variables on the mean and variance of the non-negative rate distribution $\rho(\ln k)$. We found that environment changes the mean, μ, of $\ln k$. Composition is correlated with both the mean and variance, σ^2, of $\ln k$ and therefore affects the faster rates of the distribution. We suggest that climatic and compositional variable may be similarly related to $v(\ln k)$. To evaluate that hypothesis, we proceed to compare the general mass-constrained exit rate function $v(\ln k)$ to the non-negatively constrained exit rate function $\rho(\ln k)$.

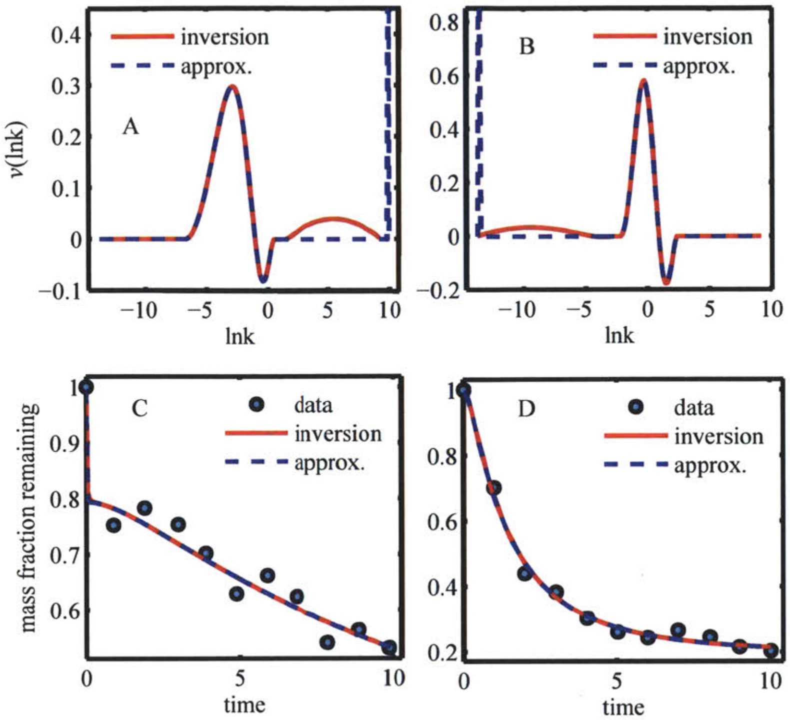

Figure 4-11: Simplification of trimodal and quadmodal datasets. A: Example of a trimodal dataset (red) and its leached fraction approximation (dashed blue). B: Example of a quadmodal dataset (red) and its approximation with an inert fraction (dashed blue). C: LIDET decay data (circles), predicted mass loss from the trimodal $v(\ln k)$ (red) and the approximation (dashed blue). D: LIDET decay data (circles), predicted mass loss from the quadmodal $v(\ln k)$ (red) and its approximation (dashed blue).

4.8 Comparing the Exit Rate Function $v(\ln k)$ to a Distribution of Rates $\rho(\ln k)$

In this section we compare the inversion results of the previous section to the inversion results when $v(\ln k)$ is constrained to be a positive distribution of log-rates $\rho(\ln k)$ [13]. Continuous distributions of decay rates have been used to model the degradation of organic matter [49, 18, 50, 39, 41, 13]. We've previously identified that the distribution associated with plant matter decay is on average lognormal [13]. Although positive rate distributions are mathematically equivalent to a continuum of parallel first order decays, they may also represent the exit rate function associated with a network of possible decay states as seen in sections 4.4, 4.5.1, and 4.6. The non-negative constraint on $\rho(\ln k)$ results in a completely monotone (essentially concave-up) decay [13]. However, the non-negative constraint restricts the configurations of possible decomposition networks which may best fit the data.

We find that the distribution $\rho(\ln k)$ is typically consistent with the positive region of $v(\ln k)$. Figure 4-12A shows a comparison of a typical bimodal exit rate function to the rate distribution

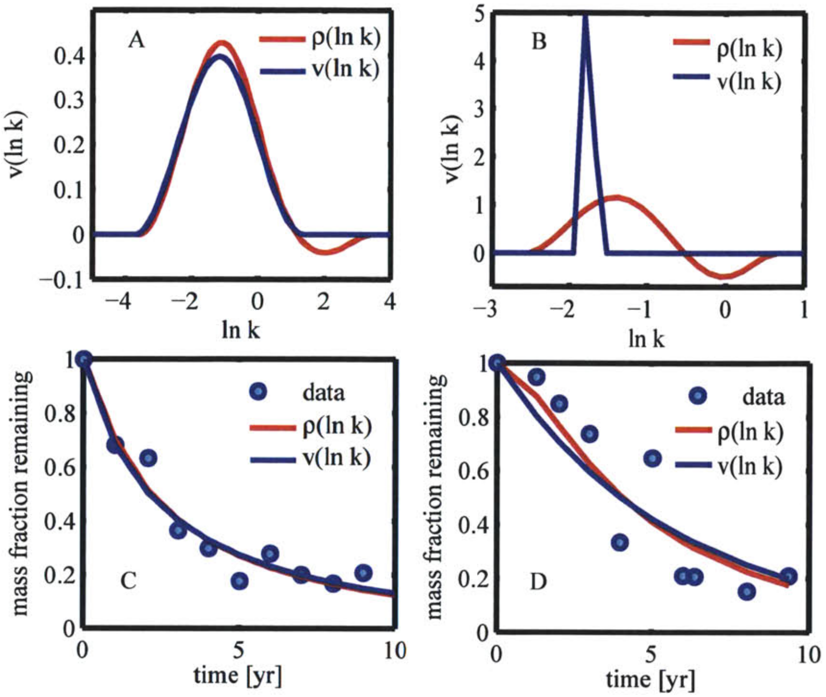

Figure 4-12: Comparison of non-negatively constrained exit rate function $\rho(\ln k)$ [13] to the mass-loss constrained exit rate function $v(\ln k)$. A: $v(\ln k)$ (red) and $\rho(\ln k)$ (blue) calculated from the inversion of the LIDET data shown in Figure C. B: $v(\ln k)$ (red) and $\rho(\ln k)$ (blue) calculated from the inversion of the LIDET data shown in Figure D. C: LIDET decay data (circles), predicted mass loss from $v(\ln k)$ (red) and $\rho(\ln k)$ (blue). D: LIDET decay data (circles), predicted mass loss from $v(\ln k)$ (red) and $\rho(\ln k)$ (blue).

$\rho(\ln k)$ resulting from the non-negative inversion. For the LIDET datasets, the negative region is typically located at fast rates and only has an effect on very early degradation times before the first measurement. The positive region explains most of the decay. Therefore a completely positive rate distribution often predicts the litter bag decay data just as well as an exit rate function with negative components as seen in Figure 4-12C. However, some data sets are more sigmoidal and exhibit a delay before decay proceeds as shown in Figure 4-12D. In this case, the negative region of the exit rate function is spread over rates associated with the timescale of the noticeable delay, whereas the best fitting $\rho(\ln k)$ is just a delta function at one rate (Figure 4-12B). Allowing for a range of serial processes therefore better captures the sigmoidal shape of the decay.

Figure 4-13 compares the mean μ_ρ and variance σ_ρ of the log rate distribution $\rho(\ln k)$ to the mean μ_{v+} and standard deviation σ_{v+} of the positive portions of the exit rate function $v(\ln k)$. We consider the 191 datasets which are well described by a distribution of decay rates [13]. To provide a relevant comparison to $\rho(\ln k)$ and in order to calculate a mean and variance, we rescale the positive portion of $v(\ln k)$ so that it integrates to 1. Figure 4-13A shows that the means μ_ρ and μ_{v+} are highly correlated with one another. Figure 4-13B also indicates that the standard deviations

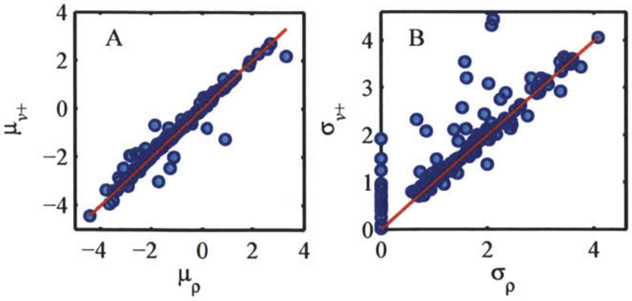

Figure 4-13: Comparison of kinetic properties of the exit rate function $v(\ln k)$ and exit rate distribution $\rho(\ln k)$. A: Plots the mean log rate μ_{v+} associated with $v(\ln k)$ against the mean log rate μ_ρ of the distribution $\rho(\ln k)$. B: Plots the standard deviation of log rates σ_{v+} associated with $v(\ln k)$ against the standard deviation σ_ρ of the distribution $\rho(\ln k)$. Red lines have a slope of 1 and intersects the origin. 191 datasets.

of the two distributions are also highly correlated. The spread in σ_{v+} at $\sigma_\rho = 0$ is associated with datasets similar to Figures 4-12B. These datasets are best fit by a single rate when the exit rate function is constrained to be non-negative, but better fit by an exit rate function resulting from a more general network of transformations. The additional outliers $\sigma_{v+} > \sigma_\rho$ are due to the presence of multiple positive modes in $v(\ln k)$ but not $\rho(\ln k)$.

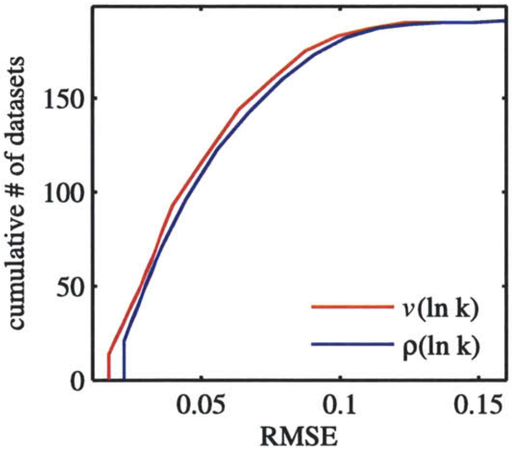

Figure 4-14: Root Mean Square Error (RMSE) of the decay predicted by $v(\ln k)$ (red) and $\rho(\ln k)$ (blue). Vertical axis indicates the cumulative number of datasets having an RMSE less than the value shown on the horizontal axis. 191 datasets

In Figure 4-14 we compare the root mean square error (RMSE) of the decay predicted from $\rho(\ln k)$ and $v(\ln k)$ for 191 LIDET datasets. Figure 4-14 plots the cumulative number of datasets having a RMSE less than the value shown on the horizontal axis. The general exit rate function $v(\ln k)$, shown in red, fits the data slightly better than the positively constrained $\rho(\ln k)$ shown in blue. The mean RMSE associated with q and ρ are 0.053 and 0.056. For comparison, the mean

RMSE of the best fitting multi-pool models is 0.050, however these solutions are highly sensitive to noise [51]. The fits are similar because many of the LIDET datasets contain little to no noticeable lag time and the decay kinetics are mostly captured in the positive region of $v(\ln k)$.

We've previously found the distribution $\rho(\ln k)$ to be well characterized by a lognormal distribution [13]. The consistency of the calculated exit rate function $v(\ln k)$ with the calculated rate distribution $\rho(\ln k)$ over moderate and large degradation timescales suggests that the exit rates from a degradation network can be well approximated by a lognormal distribution in many cases. Furthermore, the high correlation between μ_ρ and μ_{v+}, σ_ρ and σ_{v+} suggest that the trends observed between climatic factors, organic matter composition, and kinetic parameters [13] also apply to the exit rates from a decomposition network. For example, temperature and moisture appear to affect μ_ρ but not σ_ρ and thus scale all decay rates by the same factor. Furthermore, the exit rates are closely associated with the transformation rates in the decay network we can say that climatic factors like temperature and moisture effectively affect the rates of all transformation processes.

4.9 Discussion

These results suggest interpreting the continuum formulation (4.22) of organic matter decay not as a distribution of components decaying in parallel, but rather as a distribution of Poisson exit rates from an underlying decay system. Carbon in any pathway of the network may react and exit at random rates.

The rates k in the exit rate function represent the eigenvalues or poles of the underlying decay system. The exit rate function $r(k)$ or $v(\ln k)$ encodes the impulse response of the decay system. In this sense $v(\ln k)$ is related to the transfer function of the underlying decay system. The transfer function, $T(s)$, completely describes the relationship between the input and output of linear, time-invariant systems. It can be calculated by taking the Laplace transform of either the standard state-space or matrix formulation of the decomposition network (4.12) or the impulse response (4.18), yielding

$$T(s) = \mathbf{1}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{p}_0 \tag{4.28}$$

as derived in Appendix 4.10.8. Equation (4.28) shows that the poles of the system are also the eigenvalues of $\mathbf{A}$. For a continuous $r(k)$, the transfer function is

$$T(s) = \int_0^\infty e^{-st} \int_0^\infty r(k)e^{-kt}dkdt \tag{4.29}$$

$$\tag{4.30}$$

The transfer function is thus the Laplace transform of the Laplace transform of $r(k)$. Equation (4.29) shows a direct relationship between the transfer function, the impulse response, and $r(k)$ (or

$v(\ln k)$). Knowing any of these three quantities is sufficient to describe the total mass dynamics in the system. Because the underlying decay network associated with a given impulse response or $v(\ln k)$ may not be unique, the exit rate function $v(\ln k)$ provides an accurate, general, compact, and concise description of the underlying complex and unknown decomposition system.

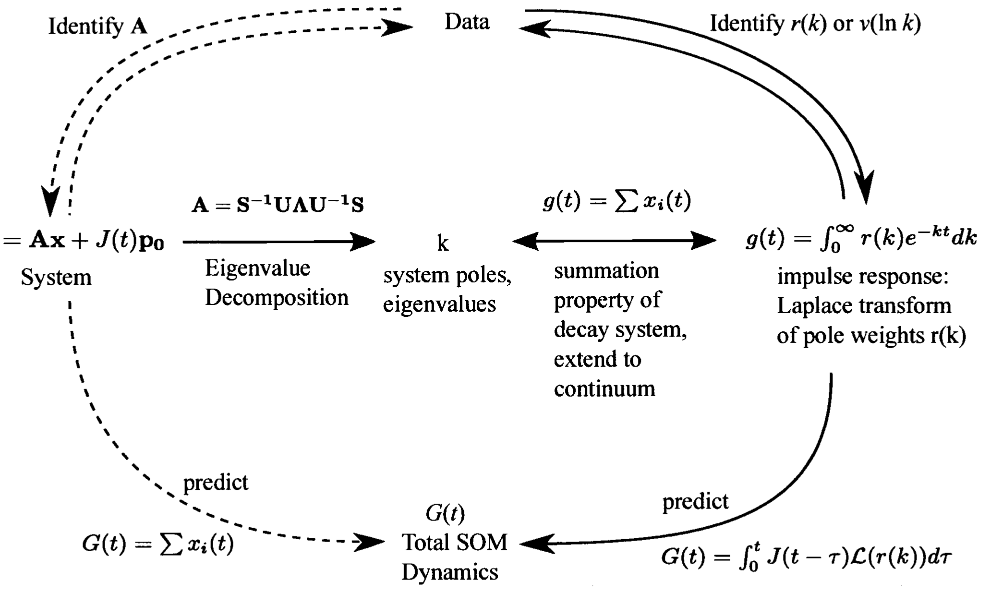

Figure 4-15: Summary of the exit rate function approach. Standard approaches for tuning decay models are shown by the dashed lines. Full network information is necessary in the standard approach to predict SOM dynamics. By writing the impulse response of the system as $v(\ln k)$ full network information is no longer necessary. Instead only the simpler exit rate function needs to be identified in order to predict SOM dynamics.

A summary of the use and interpretation of the exit rate function approach is shown in Figure 4-15. Figure 4-15 also compares the exit rate approach to standard compartmental approaches for fitting and predicting decay data, shown as the dashed lines. Compartmental approaches require tuning parameters associated with the compartments in order to fit decay data. The tuned compartmental network is then directly used to predict soil organic matter dynamics. As discussed throughout the paper, the exit rate approach is an equivalent representation of compartmental transformation models and continuum transformation models [10] with linear transformations. We advocate both system estimation and the prediction of total soil organic matter (SOM) dynamics via this simpler formulation of a heterogeneous decay network, $v(\ln k)$

We find that inversions $v(\ln k)$ of the LIDET data typically have a mostly lognormal shape with a smaller region of faster negative components. These $v(\ln k)$ may derive from a network containing two or more heterogeneous pools. More likely however, these exit rate functions derive from a continuum of states which transform at a continuum of rates. Although the inversion is general, allowing for a network of transformation processes, we often find that the dynamics of carbon degradation through these complex networks are well approximated by a parallel system.

The $v(\ln k)$ obtained from litterbag studies however only characterizes the portion of the decay network associated with young soil organic matter transformations. More work is necessary in order to probe the entire SOM network and the full domain of the exit rate function $v(\ln k)$.

Parallel, serial, and feedback processes are somewhat analogously represented in the distribution $v(\ln k)$. The presence of a negative $v(\ln k)$ indicates the existence of serial or feedback processes, while positive $v(\ln k)$ could represent parallel processes or serial and feedback processes with mass loss. In this manner identifying $v(\ln k)$ could be used as a coarse investigative tool.

Bolker et. al. [18] have questioned the advantages of the diagonalized, eigenvalue approach over the standard forward compartmental approach in soil degradation. We believe the exit rate analysis via the inverse Laplace transform is powerful tool for analyzing degradation, as it accounts for particle heterogeneity and predicts overall mass dynamics, but does not require the details of all transformations. By describing degradation at an appropriate level of model complexity, this approach provides a clear visualization of the mass dynamics of the system, as well as a simple framework for analyzing the influence of biogeochemical and environmental factors on respiration.

This approach assumes that the non-linearities in soil organic matter decay time series arise from the kinetic heterogeneity of an underlying linear decay system. Different modeling approaches may also incorporate non-linear processes [9]. These may include for example, Michaelis-Mentin enzyme kinetics [7], Lotka-Volterra decomposer growth [84], sorption and other chemical/biochemical processes [147]. However it is not clear however that enzyme kinetics, decomposer abundance, and/or sorption are the rate limiting decomposition processes. The non-linearities of other transformation processes are currently not well characterized. Thus it is difficult to accurately include nonlinear effects in large-scale degradation models.

Overall, this approach is general for tracking the flow of mass through any network approximated by linear kinetics. These methods may apply for example in molecular biology to analyze concentrations of proteins which denature and degrade at varied frequencies during various cellular processes, in pharamacokinetics [88, 89, 139, 135, 90] to characterize heterogeneous absorption timescales and predict the concentration of a drug remaining in the body.

4.10 Appendix

4.10.1 Solution to the two-state example problem

The solution to (4.4) is readily found by assuming solutions exist of the form

$$\mathbf{x}(t) = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \mathrm{e}^{-\lambda \mathrm{t}}. \tag{4.31}$$

Substituting (4.31) for $\mathbf{x}$ in equation (4.4) results in the eigenvalue problem

$$-\lambda \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -k_1 & 0 \\ fk_1 & -k_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{4.32}$$

which, after subtracting the left hand side from both sides of (4.32) is

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -k_1 + \lambda & 0 \\ fk_1 & -k_2 + \lambda \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \tag{4.33}$$

However, there are only a few possible values of the eigenvalues λ and eigenvectors $\mathbf{u}$ which can satisfy (4.33). This occurs when the matrix is singular and its determinant is zero, providing the characteristic equation

$$(-k_1 - \lambda)(-k_2 - \lambda) = 0, \tag{4.34}$$

which has the two solutions

$$\lambda_1 = k_1 \tag{4.35}$$

$$\lambda_2 = k_2. \tag{4.36}$$

The directions of the eigenvectors $\mathbf{u_1}$ and $\mathbf{u_2}$ are found by plugging λ_1 and λ_2 into (4.33).

The solution $\mathbf{x}$ is a superposition of both exponential decays,

$$\mathbf{x} = \alpha_1 \mathbf{u_1} \mathrm{e}^{-\lambda_1 \mathrm{t}} + \alpha_2 \mathbf{u_2} \mathrm{e}^{-\lambda_2 \mathrm{t}} \tag{4.37}$$

which is

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \alpha_1 \begin{bmatrix} k_2 - k_1 \\ fk_1 \end{bmatrix} \mathrm{e}^{-\lambda_1 \mathrm{t}} + \alpha_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathrm{e}^{-\lambda_2 \mathrm{t}} \tag{4.38}$$

The weights α for each eigenvector are determined by substituting the initial condition $\mathbf{x}(0)$ into (4.37) at $t = 0$, resulting in the system of equations

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = [\mathbf{u_1} \quad \mathbf{u_2}] \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}. \tag{4.39}$$

The solution to (4.39) is

$$\alpha_1 = \frac{fk_1}{k_2 - k_1} x_1(0) \tag{4.40}$$

$$\alpha_2 = x_2(0) - \frac{fk_1}{k_2 - k_1} x_1(0) \tag{4.41}$$

The total mass of the system, $G(t) = x_1(t) + x_2(t)$ therefore decays as

$$G = r_1 \mathrm{e}^{-\lambda_1 \mathrm{t}} + \mathrm{r}_2 \mathrm{e}^{-\lambda_2 \mathrm{t}} \tag{4.42}$$

where, as seen from (4.38)

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} \sum_i u_{1,i} & 0 \\ 0 & \sum_i u_{2,i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix}. \tag{4.43}$$

Solving (4.43) provides the estimates of the mass fractions r_1 and r_2 in eigenspace

$$r_1 = x_1(0) \left(1 + \frac{fk_1}{k_2 - k_1}\right), \tag{4.44}$$

$$r_2 = x_2(0) - x_1(0) \frac{fk_1}{k_2 - k_1}. \tag{4.45}$$

4.10.2 Solving a linear system

Linear problems such as (4.12) are often solved by assuming a solution of the form $\boldsymbol{u}e^{-\lambda t}$ and plugging it into (4.12), resulting in a "characteristic equation" for the eigenvalues λ. The rates are then identified by substituting those values of k back into (4.12) to identify the eigenvectors $\boldsymbol{u}$. This approach was shown in equations (4.4) - (4.40).

We use a similar eigenvalue analysis to solve (4.12). First, we decompose $\mathbf{A}$ into its eigenvalues λ_i and eigenvectors $\boldsymbol{u_i}$

$$\mathbf{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{-1} \tag{4.46}$$

where

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_z \end{bmatrix} \tag{4.47}$$

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_z \end{bmatrix} \tag{4.48}$$

Substituting (4.46) into (4.12), we get

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{x} + J(t)\mathbf{p_0} \tag{4.49}$$

Next, we change the state coordinate system x_i to one aligned with the eigenvectors. The system then becomes diagonal or parallel, in eigenspace. This is done by multiplying both sides of (4.12) by $\mathbf{U^{-1}}$,

$$\mathbf{U^{-1}}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{x} + J(t)\mathbf{U}^{-1}\mathbf{p_0} \tag{4.50}$$

and then changing the coordinate system via the transformation

$$\boldsymbol{\alpha} = \mathbf{U}^{-1}\mathbf{x}. \tag{4.51}$$

Applying the transformation (4.51) to (4.50) yields the diagonalized system

$$\frac{\mathrm{d}\boldsymbol{\alpha}}{\mathrm{d}t} = \mathbf{\Lambda}\boldsymbol{\alpha} + J(t)\mathbf{U}^{-1}\mathbf{p}_0. \tag{4.52}$$

When the forcing $J(t)$ is an impulse of mass G_0, this input becomes equivalent to letting the system decay from the initial condition $\boldsymbol{\alpha}(0) = G_0\mathbf{U}^{-1}\mathbf{p}_0$, as described in appendix 4.10.4. Thus, for an initial impulse of size G_0 system can be rewritten as

$$\frac{\mathrm{d}\boldsymbol{\alpha}}{\mathrm{d}t} = \mathbf{\Lambda}\boldsymbol{\alpha}, \quad \boldsymbol{\alpha}(0) = G_0\mathbf{U}^{-1}\mathbf{p}_0, \tag{4.53}$$

which has the solution

$$\alpha_i(t) = \alpha_i(0)e^{-\lambda_i t}. \tag{4.54}$$

or in matrix form

$$\boldsymbol{\alpha}(t) = e^{-\mathbf{\Lambda}t}\boldsymbol{\alpha}(0) = G_0 e^{-\mathbf{\Lambda}t}\mathbf{U}^{-1}\mathbf{p}_0. \tag{4.55}$$

The state dynamics $\mathbf{x}$ are then recovered by the eigenvector transformation $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha}$,

$$\mathbf{x}(t) = G_0\mathbf{U}e^{-\mathbf{\Lambda}t}\mathbf{U}^{-1}\mathbf{p}_0. \tag{4.56}$$

4.10.3 Expressing mass loss in a network as a sum of exponentials

Here, we seek to express the total mass in a network as a sum of exponential decays. The mass of organic matter G remaining in the system at time t is found by summing all states $x_i(t)$

$$G(t) = \sum_i x_i(t). \tag{4.57}$$

We now project the system into the eigen-decay state co-ordinates in order to express the mass as the standard sum of exponential decays. Substituting $\mathbf{x} = \mathbf{U}\boldsymbol{\alpha}$ into (4.57) gives

$$G(t) = \sum_i \alpha_i(t) \sum_j u_{i,j} \equiv \sum_i m_i(t) \tag{4.58}$$

The effective mass m_i in the i^{th} eigen-decay state is the weight α_i multiplied by the sum of all components of the eigenvector $\mathbf{u}_i$,

$$\mathbf{m} = \mathbf{S}\boldsymbol{\alpha} \tag{4.59}$$

where

$$\mathbf{S} = \begin{bmatrix} \sum_j u_{1,j} & 0 & \dots & 0 \\ 0 & \sum_j u_{2,j} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sum_j u_{n,j}. \end{bmatrix} \tag{4.60}$$

The relation between decaying parallel states $m_i(t)$ and the actual states $x_i(t)$ is therefore

$$\mathbf{m}(t) = \mathbf{S}\mathbf{U}^{-1}\mathbf{x}(t). \tag{4.61}$$

The matrix S simply rescales each eigenvector $\mathbf{u}_i$ by its length $\sum_j u_{i,j}$. Therefore we see that by projecting the system onto eigenvectors with unit length $\sum_j u_{i,j} = 1$, the total mass of the eigenstates represents the total system mass, $\sum_i x_i = \sum_i m_i$.

Multiplying both sides of (4.52) by $\mathbf{S}$, we find

$$\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t} = \mathbf{\Lambda}\mathbf{m} + J(t)\mathbf{r}, \tag{4.62}$$

where

$$\mathbf{r} = \mathbf{S}\boldsymbol{\alpha}(0) = \mathbf{S}\mathbf{U}^{-1}\mathbf{p}_0 \tag{4.63}$$

is the initial fractionation of incoming organic matter into all eigen-decay states. Equation (4.63) is the full version of equation (4.43) from the 2x2 system.

Because Λ is diagonal, equation (4.62) represents a parallel system of exponential decays. The solution to this system for the case of an initial input or impulse of organic matter G_0 at $t = 0$ is

$$m_i(t) = G_0 r_i e^{-\lambda_i t}, \tag{4.64}$$

as derived in Appendix 4.10.4. The total mass remaining in the system is found by summing all of the eigenstates $m_i(t)$. The fraction, $g(t) = G(t)/G_0$, of original mass remaining is

$$\begin{aligned} g(t) &= \sum_i m_i(t)/G_0 && (4.65) \\ &= \sum_i r_i e^{-\lambda_i t}. && (4.66) \end{aligned}$$

4.10.4 Impulse response

When a system receives an input at $t = 0$ and no input for $t > 0$, the response of the system to this input or impulse is called the "impulse response." An impulse G_0 of mass at $t = 0$ is modeled by setting $J(t) = G_0\delta(t)$. The immediate response of the system (4.62) is calculated by integrating both sides of (4.62) in time over the infinitesimally small duration of the impulse.

$$\int_0^{0^+} \frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t}\mathrm{d}t = \int_0^{0^+} \boldsymbol{\Lambda}\mathbf{m}\mathrm{d}t \int_0^{0^+} G_0\delta(t)\mathbf{r}\mathrm{d}t \tag{4.67}$$

Because $\mathbf{m}$ is initially the zero vector and remains finite during and after the impulse, the first term on the right hand side of (4.67) vanishes, leaving only the other two terms. Integrating both remaining terms gives

$$\mathbf{m}(0^+) - \mathbf{0} = G_0\mathbf{r} \tag{4.68}$$

thus providing the initial condition

$$\mathbf{m}(0) = G_0\mathbf{r} = G_0\mathbf{S}\mathbf{U}^{-1}\mathbf{p}_0 \tag{4.69}$$

Because $J(t) = 0$ for $t > 0$, the solution to (4.62) with the initial condition given by (4.69) is

$$m_i(t) = G_0 r_i e^{-\lambda_i t}. \tag{4.70}$$

4.10.5 Extrapolating the network to the continuum

We extrapolate the network model to the continuum by letting the number of organic matter states x_i approach infinity. The resulting continuous system is

$$\frac{d}{dt}c(x,t) = \int_0^\infty D(x,x')K(x')c(x',t)\mathrm{d}x' - d(x)c(x,t) + J(t)p_0(x) \tag{4.71}$$

The concentration vector $\mathbf{x}$ then becomes the continuous function $c(x)$. Dynamics expressed via the product $\mathbf{Ax}$ become the convolution of the the bi-variate transformation function $D(x,x')$ with the rate function $K(x',t)$ and $c(x',t)$. The transformation function $D(x,x')$ maps the transformations from states x' to x in the same way that the transition probability matrix maps transformations in a Markov chain. The rate of transformation associated out of state x' is $K(x',t)$, like the rates k_i in the network model. In this formulation I restrict $\int_0^\infty D(x,x')dx = 1$ for all x'. Because $D(x,x')$ integrates to unity, the mass flow to the inorganic (exit) state is given separately by the rate $d(x)$ which is related to the rates k_i and conversion fractions f_{ij} in $\mathbf{A}$.

With certain assumptions, the general form of the continuous q-theory [10, 144] of Agren and Bosatta describes the continuum limit of degradation networks. When assuming time-invariance and first order growth and transformation rates, the general q-theory model (equations (4.1) and (4.2) of [10]) results in (4.71). The qualities q [10] in the q-model are similar to the states x here except we allow transformations to be both biotic and abiotic. Here, organic matter states x reflect varying degrees of quality, availability, and accessibility.

Extending states with arbitrary non-exponential residence times to the continuum results in the more general system,

$$\frac{d}{dt}c(x,t) = \int_0^\infty D(x,x')K_s(x',t,c)\mathrm{d}x' - d(x,t,c) + J(t)p_0(x) \tag{4.72}$$

where transformation fluxes $K_s(x',t,c)$ from states x' are no longer restricted to be first-order exponential. While this allows for transformations with non-linear transit times, the system remains linear and time-invariant. The network model (4.12) can of course be adapted to allow changes in the network as a function of time, but then the exit rate function approach is no longer valid and closed form expressions for $g(t)$ not likely to be found. Because (4.72) remains linear, higher order transformations, such as second order growth rates, can also not be described.

However for linear, time-invariant processes, we can interpret the exit rate function $v(\ln k)$ as the projection of the initial state concentrations $c(x,0)$ onto the eigenfunctions of the integro-differential equation (4.72) of the decay continuum.

A quasi-continuum network which contains a large number of states is nevertheless an intuitive and powerful way to interpret and visualize the way that decomposers, substrate, and environment interact and the effects they have on the organic matter degradation process. This picture is somewhat clouded when considering only continuum theory.

4.10.6 Calculating $r(k)$ from a network with heterogeneous states and distributed rates

To numerically determine $r(k)$ associated with a known network, we use a Monte-Carlo simulation with N trials in order to estimate a mean $r(k)$. The procedure is as follows:

- For each trial, each pool i is fractionated into sub-states j as shown in Figure 4-2. For these simulations we divide the pools into a total number of n sub-states. The rates k_{ij} and/or the conversion fractions f_{ij} associated with each state are randomly assigned. Each state is assigned an equal partitioning $p_{ij} = 1/n$ because the stochastic information is contained in the random assignment of k_{ij}.
- For each trial, ℓ, the matrix A_ℓ is generated based on the values of k_{ij}, p_{ij}, and f_{ij} for each state and each connection.
- We then calculate the eigenvalues and eigenvectors of $\mathbf{A}_\ell$ and use equation (4.19) to calculate the eigen-decay state partition $\mathbf{r}_\ell$ for each trial.
- The previous three steps are repeated N times to generate an ensemble of initial eigen-state vectors $\mathbf{r}_\ell$.
- After removing outliers containing degenerate eigenvalues, we divide the $\ln k$ axis into equally spaced bins of width $\Delta \ln k$ and sum the weights r_i in each bin from all trials ℓ. Because the bins are logarithmically-spaced, we estimate the exit rate function in log-rate space $v(\ln k)$. $v(\ln k)$ is related to $r(k)$ by $r(k) = v(\ln k) d \ln k/dk$.
- We then divide the sum by the factor $N\Delta \ln k$ to provide the estimate

$$\mathbf{v} = \frac{\sum_{\ell=1}^{N} \mathbf{r}_\ell}{N\Delta \ln k}. \tag{4.73}$$

$\mathbf{q}$ is a discrete approximation of the continuous function $v(\ln k)$.

4.10.7 Proof that constraints on q are sufficient to guarantee $dg/dt <= 0$.

Here we prove that the constraint (4.27) on $v(\ln k)$ is sufficient to guarantee that $dg/dt < 0$ and $g(t) \geq 0$ for all times $t > 0$. We implement constraint (4.27) by evaluating it at a finite number of times $t_i = 1/K_i$, where $\ln K_i$ are the discretized nodes along the $\ln k$ axis. In other words, we check

$$\int_{-\infty}^{\ln K_i} k v(\ln k) \mathrm{d} \ln k > 0. \tag{4.74}$$

Because the system is a sum of exponential decays, mass goes to zero at infinity. The constraint $g(t) \geq 0$ is therefore redundant if $dg/dt < 0$ at all times t. We therefore only implement constraint (4.74) in our inversions.

The proof that (4.74) guarantees $dg/dt < 0$ proceeds by first showing that if there is initially net positive flux out of the the slowest two modes of different sign, there will always continue to be net positive flux out of those two modes. For each subsequently faster mode of different sign, we then show that if the total flux out of the system is initially positive, it will continue to be so.

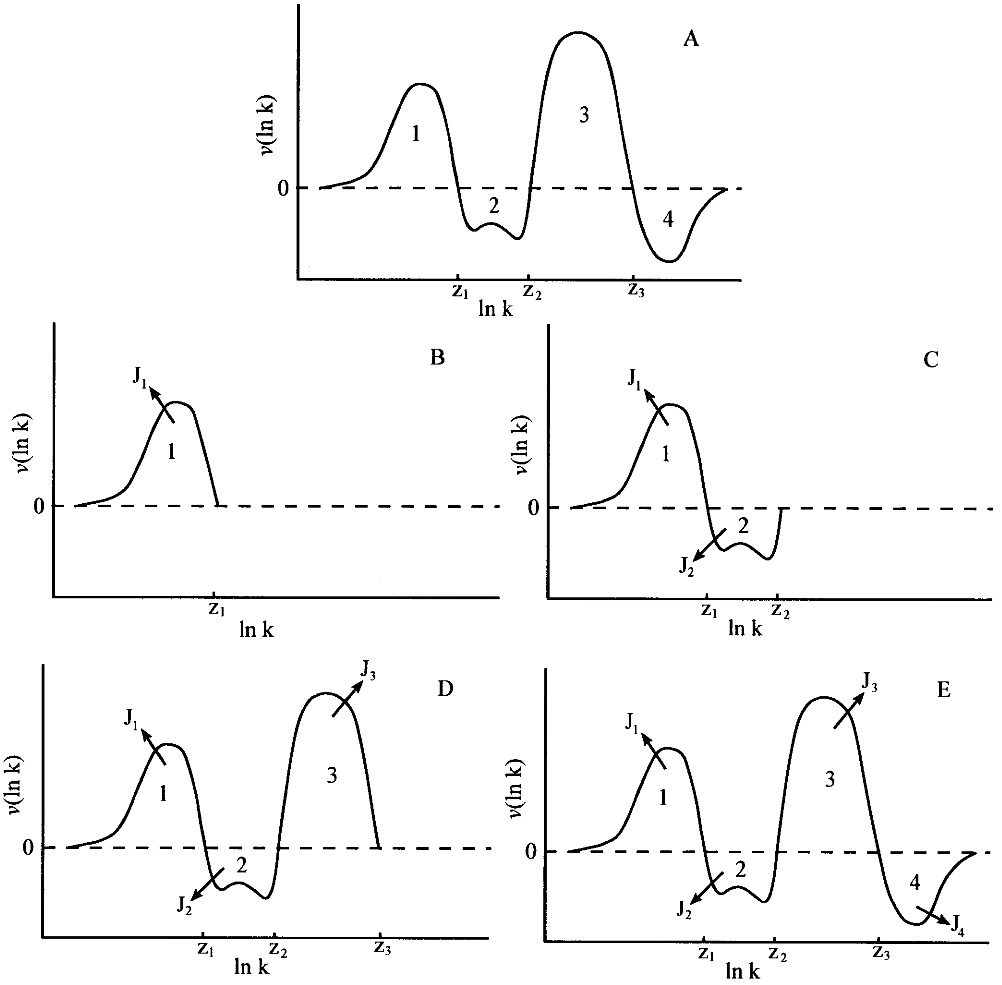

Figure 4-16: A: Arbitrary exit rate function $v(\ln k)$. Alternating positive and negative regions are marked 1-4. B: Slowest region of $v(\ln k)$. C: Regions 1 and 2 of $v(\ln k)$. D: Regions 1,2, and 3 of $v(\ln k)$. E: Regions 1,2,3, and 4 of $v(\ln k)$. In all plots, J_i indicates the flux out of region i.

Consider the arbitrary exit rate function $v(\ln k)$ in Figure 4-16A. We prove that (4.74) is sufficient to guarantee $dg(t)/dt < 0$ using the following steps:

1. For small values of K_i, the constraint (4.74) checks only the slowest section of the exit rate function. It is rather intuitive that the slowest non-zero region of the exit rate function must be positive as shown in Figure 4-16B. Otherwise there will eventually be some time t at which the mass remaining in the system will become negative. The flux out of the slowest region will always be positive because $v(\ln k)$ is positive everywhere in that domain.

2. Now consider the slowest positive and slowest negative regions of the exit rate function $v(\ln k)$, indicated as regions 1 and 2 in Figure 4-16C. We show that if the flux J_1 from positive region 1 is initially greater than the flux J_2 from negative region 2, then the total flux out of these two regions will always be positive at any time t. Specifically, we require

$$J_1(t) + J_2(t) > 0$$

or

$$J_1(t) > -J_2(t).$$

The flux J_i is calculated by taking the time derivative of the mass remaining in the i mode,

$$J_1(t) = \int_{-\infty}^{Z_1} kv(\ln k)e^{-kt}d\ln k \tag{4.75}$$

$$J_2(t) = \int_{Z_1}^{Z_2} kv(\ln k)e^{-kt}d\ln k. \tag{4.76}$$

We can rewrite $J_i(t)$ as $J_i(t) = J_i(0)j_i(t)$, where

$$j_i(t) = \frac{\int_{Z_{i-1}}^{Z_i} kv(\ln k)e^{-kt}d\ln k}{\int_{Z_{i-1}}^{Z_i} kv(\ln k)d\ln k} \tag{4.77}$$

$$\tag{4.78}$$

is simply the fraction of initial flux out of region i. It is obvious that $j_1 > j_2$ for $t > 0$ since

$$\frac{\int_{-\infty}^{Z_1} kv(\ln k)e^{-kt}d\ln k}{\int_{\infty}^{Z_1} kv(\ln k)d\ln k} > \frac{\int_{Z_1}^{Z_2} kv(\ln k)e^{-kt}d\ln k}{\int_{Z_1}^{Z_2} kv(\ln k)d\ln k}$$

and all rates associated with j_2 are greater than the rates associated with j_1.

The constraint (4.74) at $K_i = Z_i$ guarantees that $J_1(0)/-J_2(0) > 1$. Kinetics guarantee that $j_1(t)/j_2(t) > 1$. Therefore

$$\frac{J_1(t)}{-J_2(t)} = \frac{J_1(0)}{-J_2(0)}\frac{j_1(t)}{-j_2(t)} > 1. \tag{4.79}$$

3. Consider the region of the exit rate function after the next zero crossing, Z_2. This is region 3 in Figure D. We just showed that $J_1(t) > -J_2(t)$. Because region 3 must be positive, $J_3(t) > 0$ for all t. Therefore the total flux out of regions 1-3 is $J_1(t) + J_2(t) + J_3(t) > 0$.

4. Consider the 4th region of the function after the zero crossing Z_3. This mode is negative.

Using the same argument as before, the fraction of initial flux $j_4(t)$ from region 4 must follow

$$j_4 < j_3 < j_2 < j_1 \tag{4.80}$$

simply due to the kinetics of each region. The constraint (4.74) guarantees that

$$J_1(0) + J_2(0) + J_3(0) + J_4(0) > 0$$

and thus

$$\frac{J_1(0) + J_2(0) + J_3(0)}{-J_4(0)} > 1. \tag{4.81}$$

Based on the inequalities (4.80) and (4.81) we show that

$$J_1(t) + J_2(t) + J_3(t) + J_4(t) > 0$$

or equivalently

$$\frac{J_1(t) + J_2(t) + J_3(t)}{-J_4(t)} > 1,$$

because

$$\frac{J_1(0)j_1 + J_2(0)j_2 + J_3(0)j_3}{-J_4(0)j_4} > \frac{J_1(0) + J_2(0) + J_3(0)}{J_4(0)}\frac{j_3}{j_4} > 1 \tag{4.82}$$

5. Steps 3 and 4 can then be repeated for each successive sign change in $v(\ln k)$ until the final zero intersection of the function $v(\ln k)$. This guarantees that flux out of the entire system, $-dg(t)/dt$ will never be negative.

The number of constraints associated with this proof is therefore the number of times that the function $v(\ln k)$ intersects zero. The constraint (4.74) is therefore redundant because it checks for positive flux each discretized node K_i on the $\ln k$ axis. However we use (4.74) because is linear and approximately checks the intersection points without requiring knowing the exact locations Z_i of the zero intersections. Therefore the constrained inversion remains linear and computationally expedient.

Furthermore, this constraint is rather conservative, as it cuts off part of the solutions space which actually have $dg/dt < 0$ for all $t > 0$. A example of a valid solution that this constraint eliminates is when there is a large negative region next to a much larger, slightly faster positive region. The positive mode may emit more flux than the slow mode for a long time until both are negligible. However, this constraint would eliminate the possibility of this solution if the flux from the large negative pool is larger than the flux from all remaining slower modes. Therefore, this constraint tends to eliminate solutions with large fluctuations that change sign over a narrow region of $\ln k$. It is thus plausible to think of this constraint as part of the regularization process, as the constraint tends to disallow solutions with large 'roughness'.

4.10.8 Transfer function of the decomposition network

The transfer function of the decomposition network can be calculated by taking the Laplace transform of either the standard state-space or matrix formulation of the decomposition network (4.12) or the impulse response (4.18),

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathcal{J}(s)\mathbf{p_0} \tag{4.83}$$

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{p}_0\mathcal{J}(s) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) \tag{4.84}$$

where $\mathbf{X}$ is the Laplace transform, $\mathscr{L}$, of $\mathbf{x}$

$$\mathbf{X}(s) = \int_0^\infty \mathbf{x}(t)e^{-st}\mathrm{d}t \tag{4.85}$$

and $\mathcal{J}(s) = \mathscr{L}J(t)$. The total mass of the system is the sum of all states $\mathbf{x}(t)$, or

$$G(t) = \mathbf{1}^T\mathbf{x}(t). \tag{4.86}$$

The Laplace transform of equation (4.86) is

$$Y(s) = \mathbf{1}^T\mathbf{X}(s). \tag{4.87}$$

where $Y(s) = \mathscr{L}G(t)$. Substituting (4.83) into (4.87) provides the dynamic response of the mass in the system to a forcing input

$$Y(s) = \mathbf{1}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{p}_0\mathcal{J}(s) + \mathbf{1}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}_0. \tag{4.88}$$

The transfer function $T(s)$ between the output and input is

$$T(s) = \frac{Y(s)}{\mathcal{J}(s)} = \mathbf{1}^T(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{p}_0. \tag{4.89}$$

Chapter 5

Effect of Natural Selection on Degradation Kinetics

During the LIDET study, litter was collected and redistributed in foreign locations to assess the effects of litter composition and environment on degradation parameters. However, litter deposited in a foreign location may be at a disadvantage compared to native litter since microbial communities may have been selected over time to efficiently decay native litter. These communities may not be as adapted to decaying foreign litter [21], thus there may be a 'homefield advantage' for litter decaying at its native location.

In this chapter I use the parameters μ and σ which best fit the decay data $g(t)$ for 191 datasets associated with chapter 3. Of the 11 different substrate types, 10 were planted at its native 'homefield' location. Each of these 10 substrate types were planted in 17-22 different locations. Each of the 10 'homefield' locations contained between 8 and 10 different substrates, all of which were among the 'homefield' substrates.

Unfortunately the 10 substrate types were only planted in 7 common locations. Rather than limiting the analysis to only a 7x7 factor matrix with only 7 homefield datapoints, I analyze the larger but partially incomplete experimental matrix, which contains all 10 homefield data points and consists of 187/191 datasets.

I proceed by using the parameters μ and σ to estimate the strength of a 'homefield advantage' and how it effects kinetic heterogeneity.

5.1 Estimating the Homefield Effect

First, the deviation, H, from the mean μ and σ at the homefield location is calculated for each substrate.

$$H_\mu^{\text{loc}} = \mu_{\text{home}} - \overline{\mu}_{\text{sub}} \tag{5.1}$$

$$H_\sigma^{\text{loc}} = \sigma_{\text{home}} - \overline{\sigma}_{\text{sub}} \tag{5.2}$$

where μ_{home} is the μ of that substrate at the home location and $\overline{\mu}_{\text{sub}}$ is the mean μ of 17-22 different locations containing that substrate. The same notation is used for σ. The deviation from the mean μ and σ of the native substrate is also calculated for each location.

$$H_\mu^{\text{sub}} = \mu_{\text{home}} - \overline{\mu}_{\text{loc}} \tag{5.3}$$

$$H_\sigma^{\text{sub}} = \sigma_{\text{home}} - \overline{\sigma}_{\text{loc}} \tag{5.4}$$

Where μ_{home} is the μ of the that location with its native substrate and $\overline{\mu}_{\text{loc}}$ is the mean μ of the 8-10 different substrates planted at that location. The same notation is used for σ.

H^{loc} and H^{sub} from (5.1) - (5.4) are shown in figure 5-1. The top figure in 5-1 seems to indicate a homefield location effect on μ for 9 of the 10 substrates since the value of μ at the native location is above the mean μ for that substrate. σ however seems to be unaffected by the homefield. The bottom figure in 5-1 however seems to indicate no homefield effect on μ or σ. This approach however, is only partially correct because we are not including the general bias of the homefield substrate or the bias of the homefield location.

5.2 Accounting for Substrate and Location Bias

H^{loc} is a misleading measure of the homefield effect since it does not account for the bias of that location. Each homefield location may be more or less favorable to degradation than the other locations. Similarly each substrate may have a different inherent lability than the other substrates. Therefore, we wouldn't expect the most refractory substrate to be above average compared to the rest at home. Rather, we may observe a small deviation from its typical lability. Just considering Figure 5-1 leads to a false positive or false negative homefield effect.

Therefore, H^{loc} must be normalized by the bias of the homefield location and similarly H^{sub} must be normalized by the bias of the homefield substrate. The bias of a location in the dataset is

$$B_\mu^{\text{loc}} = \overline{\mu}_{\text{loc}} - \bar{\mu} \tag{5.5}$$

$$B_\sigma^{\text{loc}} = \overline{\sigma}_{\text{loc}} - \bar{\sigma} \tag{5.6}$$

where $\overline{\mu}_{\text{loc}}$ is the mean μ of 10 substrates at that location and $\bar{\mu}$ is the mean μ of all 187 datasets.

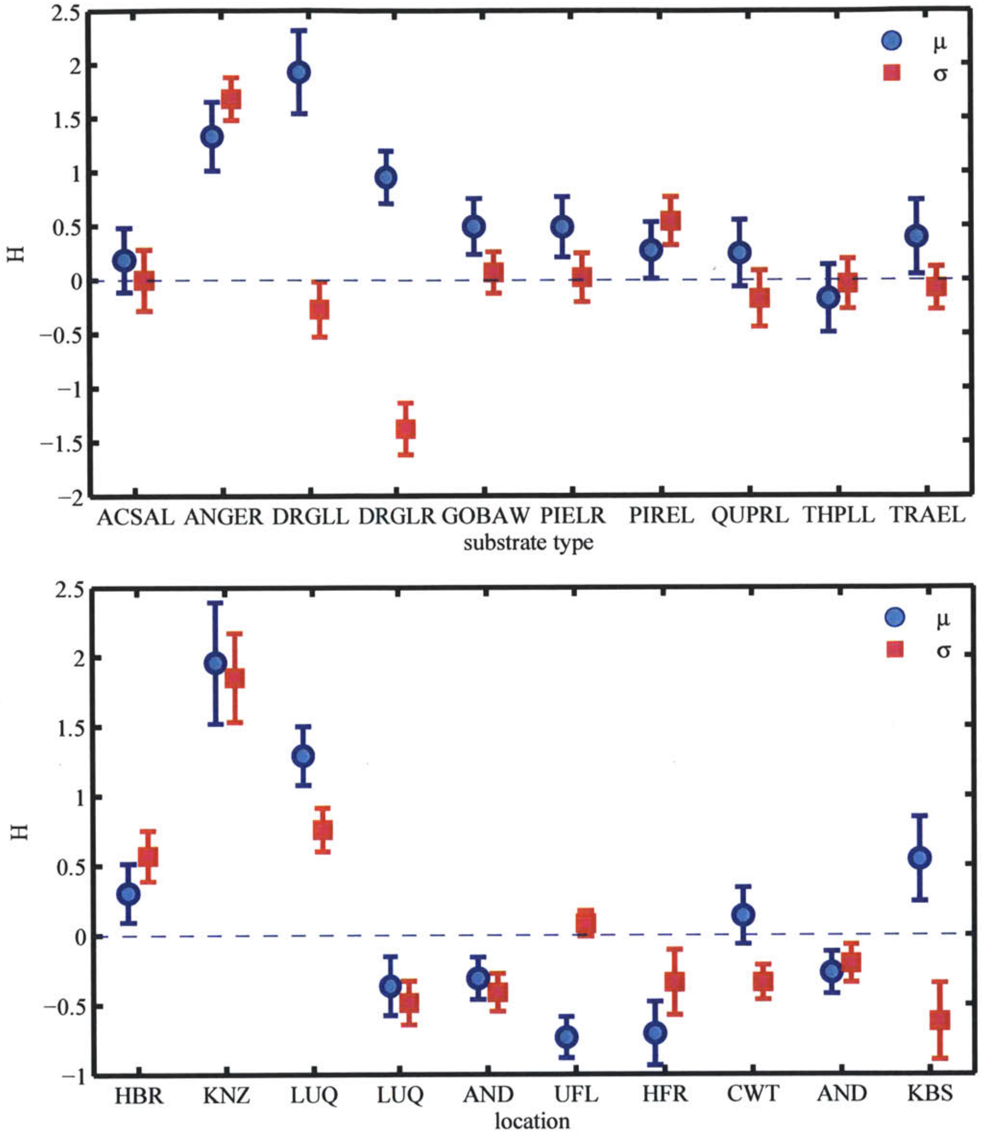

Figure 5-1: Deviation of homefield from mean values of μ and σ for the 10 native substrate/site combinations. Top: H^{loc}, the deviation of the homefield location from the mean μ and σ of 17-22 locations containing that substrate. Bottom: H^{sub}, the deviation of the native substrate from the mean of μ and σ of 8-10 substrates planted in that location. Error bars are one standard deviation of error in the estimate of $\overline{\mu}$ and $\overline{\sigma}$ Two of the substrates, DRGL (leaves) and DRGR (roots), come from the same location, LUQ.

The same notation is used for σ. Subtracting B^{loc} from from H^{loc} yields a normalized homefield effect HF

$$\text{HF}_{\mu}^{\text{loc}} = \mu_{\text{home}} - \overline{\mu}_{\text{sub}} - (\overline{\mu}_{\text{loc}} - \bar{\mu}) \tag{5.7}$$

$$\text{HF}_{\sigma}^{\text{loc}} = \sigma_{\text{home}} - \overline{\sigma}_{\text{sub}} - (\overline{\sigma}_{\text{loc}} - \bar{\sigma}) \tag{5.8}$$

Similarly the normalized homefield effect of a substrate is

$$\mathrm{HF}_{\mu}^{\mathrm{sub}} = \mu_{\mathrm{home}} - \overline{\mu}_{\mathrm{loc}} - (\overline{\mu}_{\mathrm{sub}} - \bar{\mu}) \tag{5.9}$$

$$\mathrm{HF}_{\sigma}^{\mathrm{sub}} = \sigma_{\mathrm{home}} - \overline{\sigma}_{\mathrm{loc}} - (\overline{\sigma}_{\mathrm{sub}} - \bar{\sigma}) \tag{5.10}$$

Note however that since the homefield effect has been normalized, comparing (5.7) - (5.10) reveals

$$\mathrm{HF}_{\mu}^{\mathrm{sub}} = \mathrm{HF}_{\mu}^{\mathrm{loc}} = \mathrm{HF}_{\mu} \tag{5.11}$$

$$\mathrm{HF}_{\sigma}^{\mathrm{sub}} = \mathrm{HF}_{\sigma}^{\mathrm{loc}} = \mathrm{HF}_{\sigma} \tag{5.12}$$

Since $\mu = \log k^*$,

$$\mathrm{HF}_{\mu} = \log\left(\frac{k^*}{\overline{k^*_{sub}}} / \frac{\overline{k^*_{loc}}}{\bar{k}^*}\right) \tag{5.13}$$

where $\bar{k}^*$ is a geometric mean of k^* for all of the datasets, $\overline{k^*_{sub}}$ is a geometric mean of k^* for a given substrate and $\overline{k^*_{sub}}$ is a geometric mean of k^* for a given location. HF_{μ} *is now seen as the log of the factor by which rates increase when litter degrades in its home location.*

5.3 Significance of a Homefield Effect

A plot of HF_{μ} and HF_{σ} for each of the ten homefield substrate location pairs is shown in Figure 5-2. Eight of the substate-location pairs show a positive effect of homefield on μ while only six show a positive effect on σ. The mean of HF_{μ} is 0.25 while the mean of HF_{σ} is 0.09. The P-values of each are calculated from a bootstrap technique [148] and are shown in Figure 5-3.

Figure 5-3 shows the likelihood of the null hypothesis that the mean homefield effect from all 10 substrate-location pairs came from average μ and σ of each substrate and location. The likelihood was calculated by using a bootstrap of the random variables in Eqs (5.7)- (5.13). Figure 5-3 left suggests the homefield effect HF_{μ} has a likelihood of only 0.8% of being negative (P=0.008), while the right figure suggests a likelihood of 13% that the effect on σ is negative (P=0.13). Therefore the effect $HF_{\mu} = .25$ is significant, while the effect $HF_{\sigma} = .09$ is possible but less likely. Note that I could have calculated the probability that mean homefield effect came from randomly sampling 10 substrate-location pairs from the 187 datasets (and this could have been done at the beginning of the section), but this has not much meaning since the experiment factor matrix is incomplete and there may be bias.

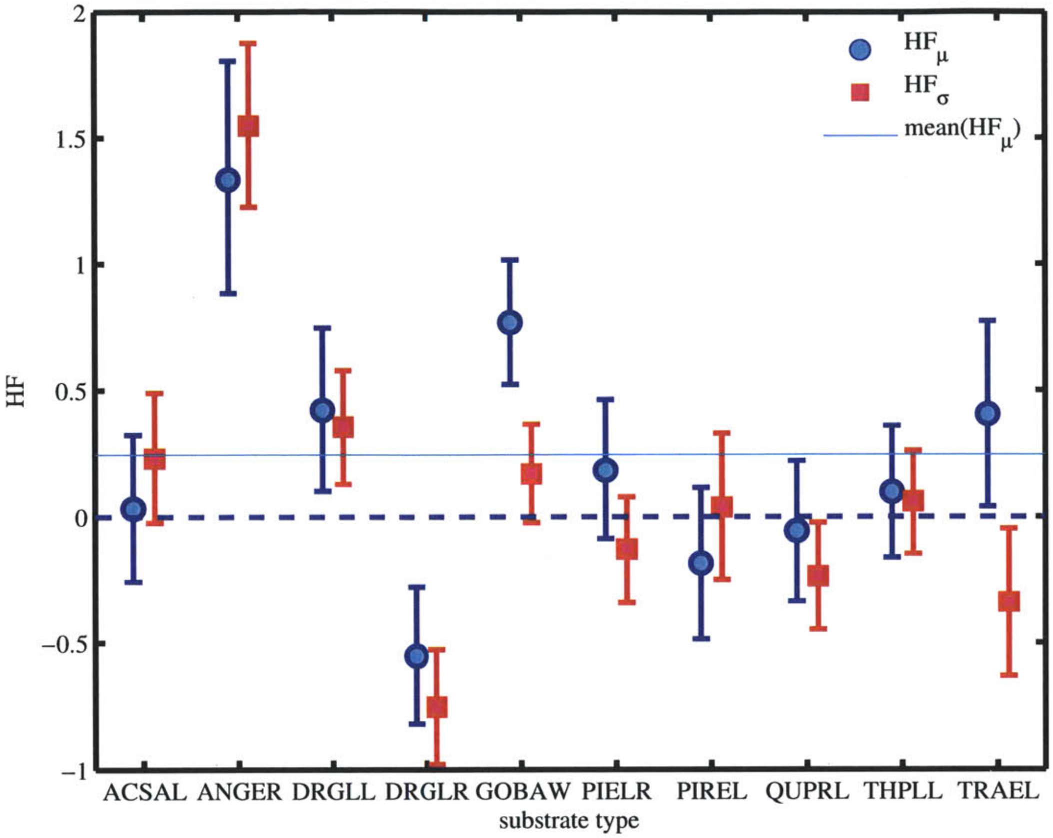

Figure 5-2: Normalized homefield effect. HF$_{mu}$ and HF$_{mu}$ is plotted for each of the 10 homefield substrate-location pairs. Error bars are based on a bootstrap calculation [148]. As shown in Figure 5-3, the mean homefield effect on μ is .28 with a P-value of 0.008. The mean homefield effect on σ is .09 with a P-value of 0.13. P-values calculated based on a null hypothesis that HF is calculated from the means of each substrate and location using a bootstrap method.

5.4 Discussion

The homefield effect on μ is to increase median rates $\ln \mu = k^*$ by a factor of $e^{.25} = 1.28$. Because changes in μ scales all rates k in the exit rate function, the homefield effect therefore appears to scale all rates by a factor of 1.28.

This finding supports the idea that local decomposer communities have been selected for efficient degradation of naturally occurring substrates. However, interpreting this effect in terms of a decomposition network yields a profound statement of the effect of natural selection on respiration kinetics in the environment. Since decomposition systems are dissipative and diagonally dominant, this behavior seems to affect all transformation rates within the decomposition network. Because many of the edge connections in a decomposition network are biotic, it seems that including the "right" species naturally selected for the particular substrate may provide additional pathways for each transformation. This may increase the mean rate of transformations in the manner depicted

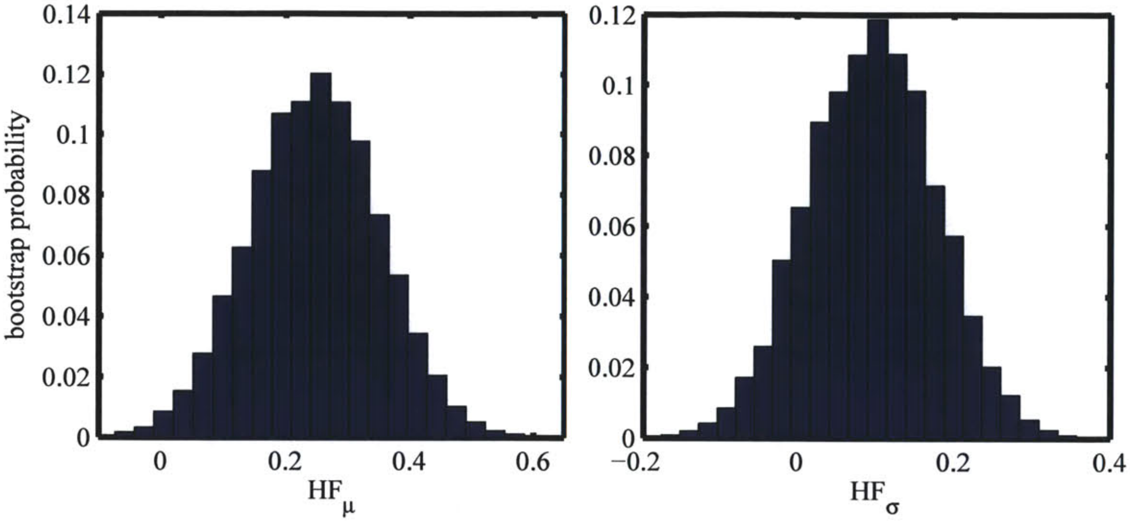

Figure 5-3: Bootstrap estimate of the likelihood that HF_μ and HF_σ could derive from the mean μ and σ of each substrate and location. HF_μ has P=0.008. HF_σ has P=0.13.

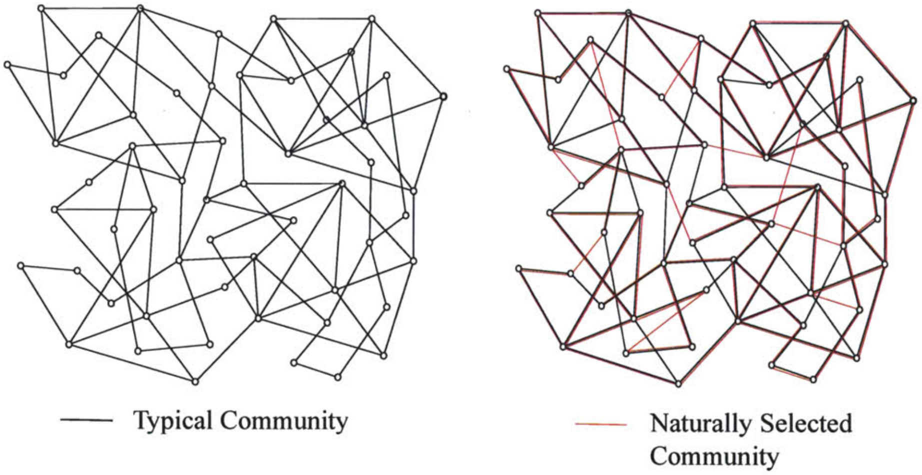

Figure 5-4: Natural selection boosts transformations throughout the network. Left: background network connectivity due to a typical decomposer community. Right: Effect of adding naturally selected species is to increase the number of pathways available for transformations and possibly forging new ones, thereby increasing all rates of exit from the network.

in Figure 1-5A. Therefore natural selection seems to boost the rates of transformation throughout the network. This effect is depicted in Figure 5-4.

With respect to the LIDET study, it means that placing litter in foreign locations has an effect on μ that reduces degradation rates by $1 - 1/1.28 = 22\%$ in addition to effects of changing temperature, precipitation and other climatic and environmental variables.

Chapter 6

Litter Decay with Climatic Fluctuations

This chapter discusses the sensitivity of decay to climatic and environmental fluctuations. The correlations between decay and climatic fluctuations are rather straightforward, as chapter 2 revealed that climatic parameters appear to be related to μ and not σ.

To address this question, I apply the distributed rate model (2.1) with lognormally distributed rates, assuming that that temperature T and other climatic variables affect μ only. I show to to predict decay in between season and under climate change. I also interpret values of μ fitted from periodic decay data from fluctuating climates.

6.1 Effect of Climate on Homogeneous Decay

Here, I consider the case of one rate that fluctuates with temperature T or time. I interpret the meaning of a measured decay rate when decay fluctuates over time, but measurements are only taken periodically at seasonal intervals.

The mass loss with the temperature fluctuations is

$$\frac{\mathrm{d}g(t)}{\mathrm{d}t} = k(t)\mathrm{d}t \tag{6.1}$$

$$\ln g(a)/g(0) = \int_0^a k(t)\mathrm{d}t \tag{6.2}$$

where $g(a)$ is mass of organic matter. At time a, the fraction of mass remaining $g(a)/g(0)$ can be described by an average mean rate $\bar{k}$ over time

$$g(a)/g(0) = e^{\int_0^a k(t)\mathrm{d}t} \tag{6.3}$$

$$g(a)/g(0) = e^{-\bar{k}a} \tag{6.4}$$

Where $\bar{k}$ is the time-averaged rate

$$\bar{k} = \int_0^a k(t)\mathrm{d}t \tag{6.5}$$

6.1.1 Periodic $T(t)$

If $T(t)$ is periodic with period $t = \tau$, then after one period

$$\begin{aligned} \int_{g((i-1)\tau)}^{g(i\tau)} \frac{\mathrm{d}g}{g} &= \int_{(i-1)\tau}^{i\tau} k(t)\mathrm{d}t && (6.6) \\ \ln \frac{g(i\tau)}{g((i-1)\tau)} &= e^{-\bar{k}\tau} && (6.7) \end{aligned}$$

The same fluctuation is repeated over the first period, second period, third and so on. Therefore, the $\bar{k}$ is able to predict decays periodically at $t = j\tau$ but not for times in between periods.

$$\begin{aligned} \ln \frac{g(j\tau)}{g(0)} &= -\sum_i^j \int_{(i-1)\tau}^{i\tau} k(t)\mathrm{d}t && (6.8) \\ &= -j\bar{k}\tau && (6.9) \\ g(j\tau) &= g(0)e^{-j\bar{k}\tau} && (6.10) \end{aligned}$$

6.1.2 Rescaling rates: averaging over seasonal variations

Using the exponential decay (6.10) to predict data between periods is effectively a prediction of decay which is averaged over seasonal phases. For instance it would describe the average decay from a larger study with 52 experiments each started at a different week in the year.

6.2 Continuum of Decay Rates

Now consider a decay composed of a distribution of first order reaction rates. From chapter 2, it seems that changes in temperature scale all rates in the continuum by the same factor. The statistical finding makes physical sense and we will assume here that all rates scale by the same factor with temperature.

Each rate $k_i(T)$ is a random variable which is multiplied by temperature factor, $R(T)$. $k_i(T(t)) = R(T(t))r_i$ where r_i is related to all other variables controlling decomposition. Each rate k_i is described in log space as

$$\log k_i(t) = \log r_i + \log R(t). \tag{6.11}$$

Eq 6.11 easily shows that changes in temperature shift all rates by the same amount in log space and therefore shifts the mean of log rates μ. Therefore the mean of log rates is temperature dependent $\mu(T)$.

6.2.1 Rescaling time

Changing the temperature can be described as rescaling time. Since the each component is affected by the same rate factor and is therefore like speeding up or retarding time. Rescaling of time and rates by $\mu(T)$: $y = te^{\mu}$ and $\kappa = ke^{-\mu}$ results in a degradation model with non-dimensional time

$$g_{\sigma}(y) = \int_0^{\infty} \Lambda(\kappa, 0, \sigma) e^{-\kappa y} d\kappa \tag{6.12}$$

where $\Lambda(\kappa, 0, \sigma)$ is a lognormal distribution of rates with log-mean of 0 and log standard deviation of σ.

6.2.2 Periodic temperature fluctuation

Since the random rate model is linear, the effect of periodic temperature fluctuation is the same for the entire continuum as it is for an individual rate as described in the previous sections. There is a weighted averaged $\bar{\mu}$ which can be used to predict decays cyclically, with cycles of one year. It is this $\bar{\mu}$ that we fit to the data since we have only yearly decay information and it is also this same $\bar{\mu}$ used in statistical analysis since we are interested in correlations with mean annual temperature information. Decay predictions within cycles is again a prediction which represents the decay averaged over all phases of the year.

6.2.3 Predicting seasonal or monthly decay

Using information about growing season and yearly temperature fluctuations, it is possible to predict seasonal mass loss using monthly temperature information and the simple relation for $\mu(T)$ from chapter 2. During the time of the year where soil is frozen, $\mu = -\infty$ and $\mu(T(t))$ can be fitted for the remaining time. Since climatic parameters such as rain or temperature only affect μ, knowledge of $\bar{\mu}(T, p)$, $T(t)$, and $p(t)$ where p here is precipitation, results in a way to predict degradation between seasons. The mean $\bar{\mu}(t)$ is the 'running average' of the parameter μ over time (based on running averages of temperature, precipitation, etc.)

$$g(t)/g(0) = \int_0^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{exp}\left(-\frac{(\ln \mathrm{k} - \bar{\mu}(\mathrm{T}(\mathrm{t}), \mathrm{p}(\mathrm{t})))^2}{2\sigma^2}\right) \mathrm{e}^{-\mathrm{kt}} \mathrm{dk} \tag{6.13}$$

Fluctuations at later times have a smaller effect as the mean signal prevails. However, only a smaller mass fraction remains at later times.

6.3 Random Temperature Fluctuations

6.3.1 Single rate

Consider an exponential decay where the average temperature over a cycle of time τ varies randomly, resulting in a random $\bar{k}$ with mean m and variance s^2. After a number of cycles n the mass remaining, $g(t)$ is

$$g(t) = e^{-\bar{k}_1\tau} e^{-\bar{k}_2\tau} e^{-\bar{k}_3\tau} e^{-\bar{k}_4\tau} ... e^{-\bar{k}_n\tau} \tag{6.14}$$

and $\log g$ is

$$\log g(t) = \tau(-\bar{k}_1 - \bar{k}_2 - \bar{k}_3 - \bar{k}_4 - ...\bar{k}_n) \tag{6.15}$$

After many cycles n, $\log g$ will be a normally distributed random variable with approximate mean of $-\tau m n$ and variance $\tau^2 s^2 n$ according to the central limit theorem. This means $g(t)$ is lognormally distributed with mean $e^{n(-\tau m+\tau^2 s^2/2)}$ and standard deviation $\sqrt{e^{n\tau^2 s^2} - 1}e^{n(-\tau m+\tau^2 s^2/2)}$.

When $(-\tau m+\tau^2 s^2/2) < 0$, which is typical, the mean and variance decrease with time, meaning random fluctuations lead to more narrowly distributed decay as degradation progresses. This appears to be a typical case for something like random variations of a rate constant with environmental temperature.

For $(-\tau m + \tau^2 s^2/2) > 0$, the distribution of decay does approach a lognormal as predicted by the central limit theorem, however predicts a mean mass remaining which increases with time as the distribution *heavily weighted by a tail which does not exist* since each $k_i > 0$.

6.3.2 Reactive continuum with random temperatures

To investigate the effect of random temperature fluctuation in a reactive continuum, we will use the property that changing μ is like rescaling time, $y = te^{\mu}$ and mass remaining is a function of y and σ only. The total rescaled time after n cycles is a random variable

$$Y = e^{\mu_1}\tau + e^{\mu_2}\tau + e^{\mu_3}\tau + e^{\mu_4}\tau + ...e^{\mu_n}\tau, \tag{6.16}$$

where $\mu_i = \bar{\mu}$ during the i^{th} cycle. The central limit theorem therefore predicts a normal distribution of Y. The distribution of mass remaining is found by substituting time with Y.

$$g(Y, 0, \sigma) = \int_0^\infty 1/\sqrt{2\pi\sigma^2} e^{-\ln k^2/(2\sigma^2)} e^{-kY} dk \tag{6.17}$$

When $\sigma = 0$, $g = e^{-Y}$ and therefore mass remaining is lognormally distributed as discussed in section 6.3.1.

6.3.3 Interpreting fitted values of μ

Here I interpret values of μ which are fitted from periodic decay data with a fluctuating climate (such as LIDET). Relations (6.16) and (6.15) also reveal that $Y = te^{\bar{\mu}}$ where $t = n\tau$. Therefore $e^{\bar{\mu}}$, it is simply the time averaged e^{μ}

$$e^{\bar{\mu}} = \frac{\int_0^{n\tau} e^{\mu(t)} dt}{n\tau}. \tag{6.18}$$

The value of μ which is fit to datasets therefore represents the average over all fluctuations in the rate $e^{\mu(t)}$

$$\mu_{\text{fit}} = \ln e^{\bar{\mu}}. \tag{6.19}$$

In the case of $\sigma = 0$, the time averaged rate k.

$$\bar{k} = \frac{\int_0^{n\tau} k(T(t)) dt}{\int_0^{n\tau} dt} \tag{6.20}$$

And thus (6.5) is equivalent to (6.20) since

$$g(a) = e^{k_1 dt + k_2 dt + k_3 dt + \ldots} \tag{6.21}$$

$$g(a) = e^{\int_0^a k(t) dt} \tag{6.22}$$

$$g(a) = e^{\bar{k}a}. \tag{6.23}$$

6.4 Effect of Increasing Temperature

To analyze the effect of an increase of temperature on the random rate model it is again useful to think of an increase in μ as speeding up time. A temperature increase which increases μ by a natural order of magnitude, e, increases Y by a factor of e since $Y = e^{\mu+1}t$. Therefore decay taking usually taking a time t, now takes place in a time $e^{-1}t$.

6.5 Interpreting the Effect of Precipitation

Precipitation has the same effect on decomposition as temperature: it changes μ but not σ, suggesting that soil moisture content affects fast and slow, in a similar manner. A lack of water should halt most biological processes, and we find this is the case. For example if a soil is dry for a fraction of the year, the effective yearly degradation timescale of all processes should be rescaled by the same factor of time, as described in the previous section. The effect of precipitation is therefore the same as temperature. In fact it seems all climatic parameters measured in the LIDET study are correlated to μ but not σ. Fluctuations in all climatic parameters can be modeled in the same

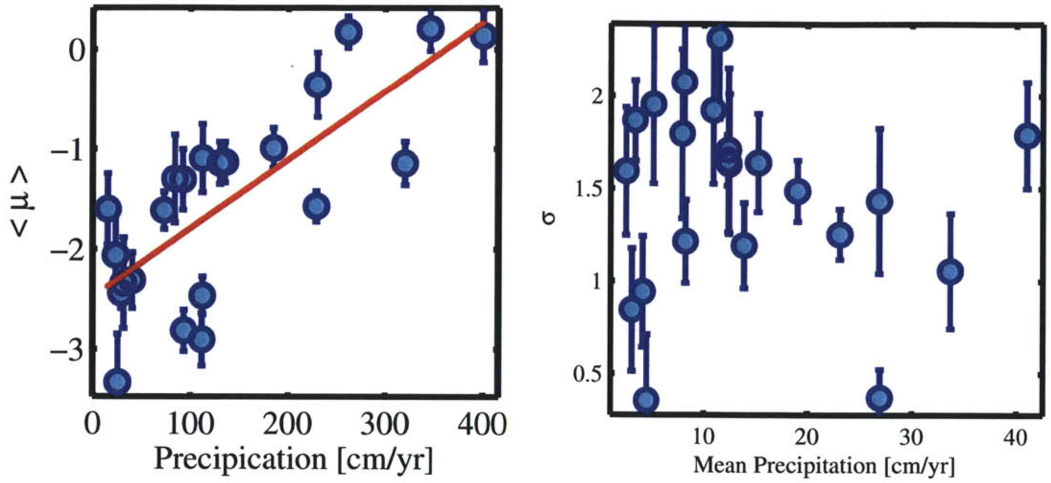

Figure 6-1: Precipitation vs. μ and σ. Precipitation is significantly correlated with μ, $r_s = .62, P = 4E-3$ but not with σ, $P = .90$. Climatic parameters such as temperature, and precipitation therefore seem to affect all rates, fast and slow.

manner discussed throughout this chapter.

6.6 Extension to the Exit Rate Function $v(\ln k)$

This approach can be extended to a general exit rate function $v(\ln k)$. Because climate affects μ only and not σ, this suggests all processes in the decomposition network equally speed up or slow down as a function of temperature and moisture. This effect on overall decay is modeled by just applying the time shift of section 6.4 to the decay $g(t)$ or shifting the entire exit rate function $v(\ln k)$ on the $\ln k$ axis.

6.7 Effect of Temperature on σ Agrees with Arrhenius

There is a slight negative correlation between temperature and σ. However, as shown in Figure 2-2C, the noise in σ is so high that this correlation is insignificant with a P-value of 0.56. Although insignificant, the observed average change in σ agrees exactly with Arrhenius kinetics

$$k = Ae^{-E_0/RT}, \tag{6.24}$$

where k is a rate, E_0 is an activation energy, R is the gas constant, T is temperature, and A is an physical encounter frequency. If the dispersion in rates k is due solely to the dispersion in the activation energy E_0 [29], then lognormally distributed rates may arise from normally distributed activation energies. According to equation (6.24), changes in temperature therefore preserve the lognormal shape of the distribution, shifting μ, and shrinking σ. Arrhenius kinetics indicate chang-

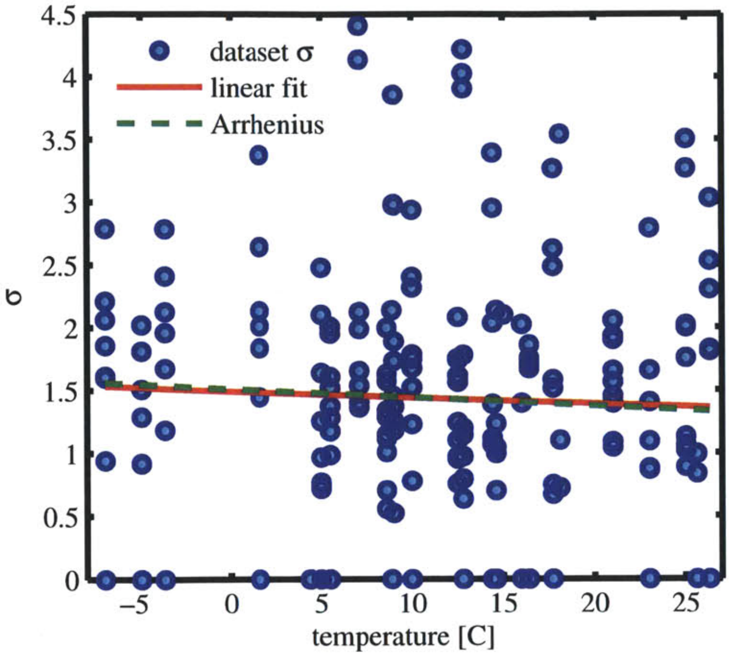

Figure 6-2: Effect of temperature on σ, as measured during the LIDET study. Red line is a linear fit of the measured σ for each dataset vs. temperature. Green dashed line corresponds to the Arrhenius prediction of how σ decreases with temperature.

ing temperature from T_0 to T_1 inversely effects the standard deviation of $\ln k$

$$\sigma_1 = \sigma_0 T_0 / T_1. \tag{6.25}$$

Figure 6-2 shows that on average across the LIDET study, σ changes with temperature in agreement with Arrhenius, although of course much noise is present. Although "biogeochemically recalcitrant organic matter will respond the most sensitively to anticipated warming" [29], this effect appears small.

The observed change in the parameter μ also agrees with both Arrhenius and $Q_{10} = 2.0$, ($d\mu/dT = \ln 0.2$, chapter 2), over the temperature range of the LIDET study.

6.8 Step Response of Carbon Stores from Changes in Climate

Because the mean ages of carbon stores in the soil may range from 10^3-10^5 [93, 37] years, changes in climate over 100's of years may be approximated by a rapid, step change. I proceed to look at the dynamic transient associated with a step change in decay parameters.

The transient response in total carbon storage $G(t)$ of a soil carbon system from a new forest (zero mass), to steady state is given by equation (A.2). The transient response for a constant input

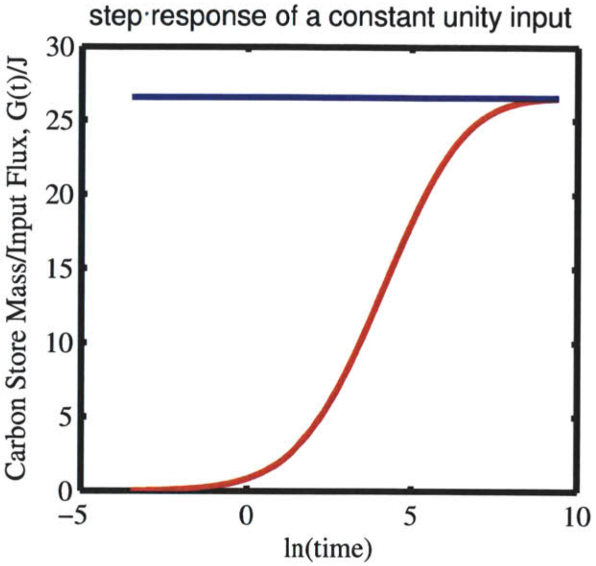

Figure 6-3: Step response from a step input from $J(t) = 0$ to $J(t) = J$ at $t = 0$. Particles have lognormal decay rates with $\mu = -2$ and $\sigma = 1.6$. Response is calculated from convolution equation and is equivalent to equation 6.28

$J(t) = J$ is

$$G(t) = \int_0^t J \int_0^\infty p(k)e^{-kt}dkdt. \tag{6.26}$$

Dividing both sides of (6.26) by J and evaluating the outer integral gives

$$G(t)/J = \int_0^\infty -k^{-1}p(k)e^{-kt}dk - \int_0^\infty -k^{-1}p(k)e^{-k0}dk \tag{6.27}$$

$$= \langle\tau\rangle - \int_0^\infty k^{-1}p(k)e^{-kt}dk. \tag{6.28}$$

Figure 6-3 shows the step response from corresponding to $J(t) = 0$ when $t < 0$ and $J(t) = J$ constant for $t \geq 0$. Particles entering the system have $\mu = -2$ and $\sigma = 1.6$. The soil system asymptotically approaches the steady state mass according to equation (6.28). The 50% rise time θ_{50} associated with the transient response (time to reach 50% of the steady state mass) is $\theta_{50} = 58.0$ yr, if the input J is given in years.

Whereas μ simply rescales all timescales (including the rise time), the effect of σ on the rise time is not as straightforward. Figure 6-4 shows the rise time as a function of σ while holding $\mu = -2$. For homogeneous carbon with only one decay rate k_1, the rise time is readily calculated from convolving the exponential impulse response with a constant input J. The 50% rise time of such a system is

$$\theta_{50} = \ln(0.5)k_1^{-1} \tag{6.29}$$

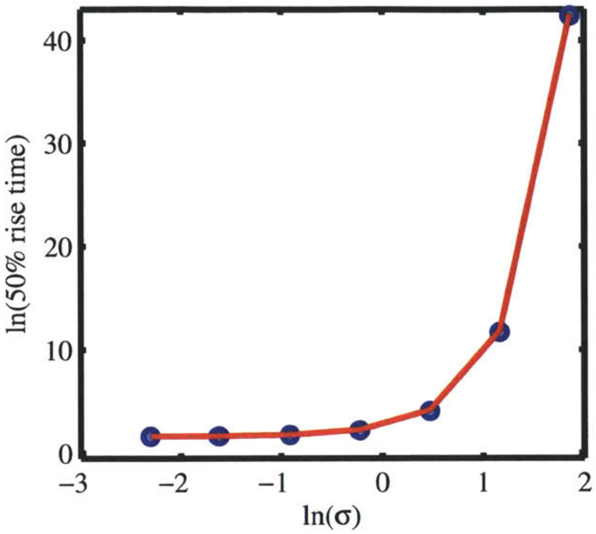

Figure 6-4: 50% rise time as a function of $\ln\sigma$. Values of σ shown here are [.1, .2, .4, .8, 1.6, 3.2, 6.4]. Data points calculated from convolution equation. Prediction, in red, is from equation 6.30.

Figure 6-4 shows that the effect of rate heterogeneity multiplies the rise time by a factor of e^{σ^2}, similar to its effect on the turnover time and the mean age. This empirical observation suggests the following form of the rise time associated with time-invariant lognormal decay:

$$\theta_{50} = \ln(0.5)e^{-\mu}e^{\sigma^2} = \ln(0.5)\tau e^{\sigma^2/2}. \tag{6.30}$$

Equation (6.30) is plotted as the red line and provides a good fit to the observed rise times. The theoretical derivation of (6.30) from (6.28) however is not straightforward and not performed for this thesis. Similarly, we show in section C.3.3 that the mean age differs from the turnover time by a factor of e^{σ^2}.

Next, I model the response from a step change in decay parameters at time $t = t_1$. I restart time at 0 from t_1. The incoming mass decays with the new parameters and new distribution p'. The old, steady mass decays with the new parameters as well but I will model its decay separately and consider the superposition of both. The previous, steady state mass was described by the steady state soil rate distribution $G_0 = \int_0^\infty Jk^{-1}p(k)dk$. Appendix A shows that the turnover time $\langle\tau\rangle$ is given by $\langle\tau\rangle = G/J$. After this disturbance the response is

$$\begin{aligned}
G(t) &= \int_0^t J \int_0^\infty p'(k)e^{-kt}dkdt + G_0(t) && (6.31)\\
&= J\int_0^\infty -k^{-1}p'(k)e^{-kt}dk - J\int_0^\infty -k^{-1}p'(k)e^{-kt}dk + \frac{G_0}{\langle\tau\rangle}\int_0^\infty k^{-1}p'(k)e^{kt}dk && (6.32)\\
&= J\langle\tau\rangle - \left(J - \frac{G_0}{\langle\tau\rangle}\right)\int_0^\infty k^{-1}p'(k)e^{-kt}dk. && (6.33)
\end{aligned}$$

The transient response from an instantaneous change in μ from $\mu = -2$ to $\mu = -1.31$ (a factor of 2, corresponding to a 10 deg C change in temperature) change is shown in Figure 6-5. The parameter σ is steady at $\sigma = 1.6$.

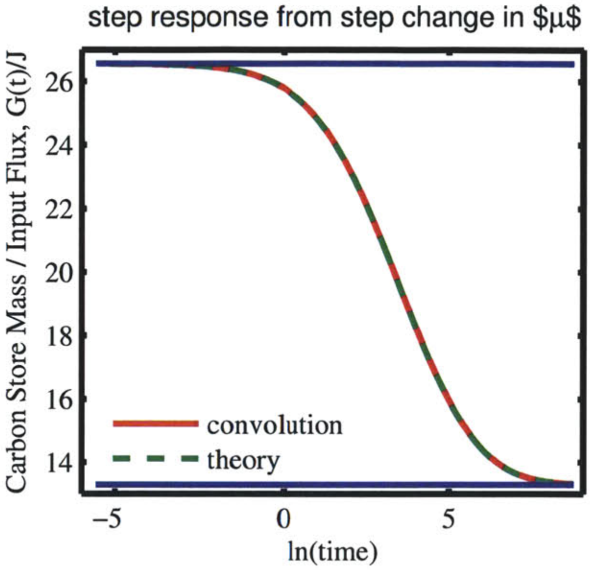

Figure 6-5: Transient response from a step change in μ, from $\mu_1 = -2.00$ to $\mu_2 = -1.31$. $\sigma = 1.6$. Red line is calculated from the convolution equation (6.31) and is in agreement with the theoretical simplification from equation (6.33) shown as dashed green line.

The rise time corresponding to the transient between two steady states is given by equation (6.30) using the value of μ_2 for the parameter μ. The 50% rise time associated with the same step change in μ for various σ is shown in Figure 6-6.

6.9 Incorporating Time-Varying Decay Systems in Carbon Cycle Simulations

If the system is not time-invariant (for example if climate is continually changing) then its difficult to model the system via the LTI convolution equation (A.2). To model changes due to climate, equation (6.13) can be used in conjunction with the time-dependent mean $\bar{\mu}(T, p, ...)$. However it is easier to track the soil distribution of rates $s(k)$ over time. There are two ways to track the soil rate distribution:

1) $s(k)$ can be discretized into small pools s_i with rates k_i. Rates k_i may vary with temperature using Q_{10}, Arrhenius, or any other preferred model. Other climatic effects may be included as well. This system is simply described as

$$\mathrm{d}s_i/\mathrm{d}t = -k_i s_i + J(t) p_i, \tag{6.34}$$

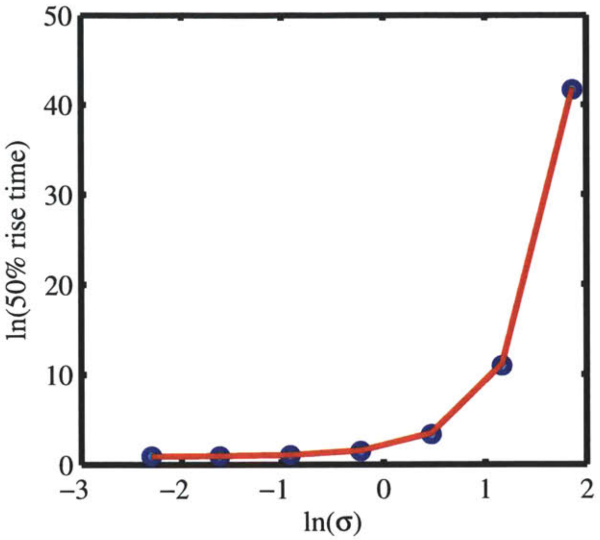

Figure 6-6: 50% rise time associated with the response due to a step change in μ plotted as a function of $\ln \sigma$. In each case $\mu_1 = -2.00$ and $\mu_2 = -1.31$ (corresponding to a factor of 2 change in all rates). The data points come from the convolution equation (6.31) and the red line is the prediction from equation (6.30).

where J is the incoming flux of mass to the system and p_i is the fraction of incoming mass with rate k_i.

2) A continuous climate advection model may be used. The dynamic equation for the soil distribution $s(k)$ can be derived from the the differential volume $s(k)\Delta k$ shown in Figure 6-7:

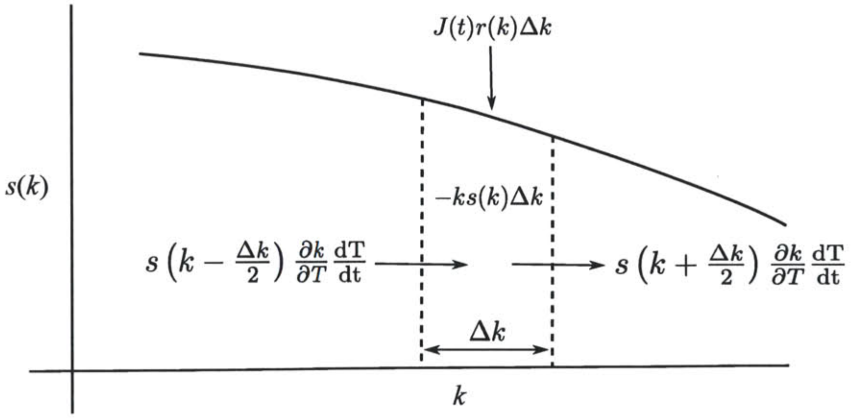

Figure 6-7: Mass dynamics of the differential volume of soil carbon $s(k)\Delta k$ at rate k. Rates k can shift with time, temperature, or other environmental parameters, resulting in advection of $s(k)$ along the k axis.

$$\frac{\mathrm{d}s(k,t)}{\mathrm{d}t}\Delta k = J(t)r(k) - ks(k) + s\left(k - \frac{\Delta k}{2}\right)\frac{\partial k}{\partial T}\frac{\mathrm{dT}}{\mathrm{dt}} - s\left(k + \frac{\Delta k}{2}\right)\frac{\partial k}{\partial t}\frac{\mathrm{dT}}{\mathrm{dt}}. \tag{6.35}$$

Dividing both sides of (6.35) by Δk, the dynamics of the soil carbon distribution can be written as the PDE

$$\frac{\mathrm{d}s(k,t)}{\mathrm{d}t} = J(t)p(k) - ks(k) - \frac{\mathrm{d}s}{\mathrm{d}k}\frac{Dk}{Dt} \tag{6.36}$$

where $\frac{Dk}{Dt}$ is the total derivative

$$\frac{Dk}{Dt} = \frac{\partial k}{\partial T}\frac{\mathrm{d}T}{\mathrm{d}t} + \frac{\partial k}{\partial c}\frac{\mathrm{d}c}{\mathrm{d}t} + \frac{\partial k}{\partial E}\frac{\mathrm{d}E}{\mathrm{d}t} + \ldots. \tag{6.37}$$

Here, c is precipitation, E is evapotranspiration, and $(\ldots)$ includes additional desired climatic or environmental effects.

Chapter 7

Conclusions

The lognormal distribution of organic matter decay rates discussed in chapter 2 is a surprisingly simple result which suggests that decomposition rates may originate from multiple stochastic requirements for decay. This origin may also relate to the observed dominant positive mode of the exit rate functions $v(\ln k)$ estimated in chapter 4. This positive mode contains most of the kinetic information regarding decay heterogeneity of the network and may be approximated as having lognormal shape. The positive constraint on $\rho(\ln k)$ discussed in chapters 2 and 3 may be interpreted as restricting the underlying decay system to a network $\mathbf{A}'$, where $\frac{d}{dt}\mathbf{x} = \mathbf{A}'\mathbf{x}$. The network $\mathbf{A}'$ is an approximation of the true network $\mathbf{A}$ where some serial connections are prohibited. The important connections however, which result in heterogeneous decay, remain. In many cases this means neglecting serial connections which are fast relative to subsequent processes in the rest of the network [18,13]. However, in the cases where a noticeable delay is apparent, this may be a bad approximation.

The rates k in the exit rate function represent the eigenvalues or poles of the underlying decay system. $v(\ln k)$ encodes the impulse response of the decay system. In this sense $v(\ln k)$ is related to the transfer function of the underlying decay system as discussed in section 4.10.8. The underlying decay network associated with the impulse response, or $v(\ln k)$, may not be unique. The exit rate function $v(\ln k)$ therefore provides an accurate, general, compact, and concise description of the underlying complex, and rather arbitrary decomposition system.

A significant contribution of this work is that it unifies many other theories and models of organic matter decomposition. Discrete compartmental transformational models [46,92,93], linear continuous transformational models [10] (see section 4.10.5), multi-pool models [22,24,42], and reactive continuum models [49,50,41] can all be expressed in terms of one another. The exit rate function $v(\ln k)$ contains the minimum information necessary to describe mass flux through a degradation network and therefore appears to be the appropriate form for analyzing decay data. From it, the decay system can be identified as discrete or continuous and the properties of possible transformations in a decay network can be evaluated. Before, models having just a few states, such as CENTURY, were seen as complex and simplification was achieved by reducing the number of states in the system [18]. The quasi-continuum transformation network and/or semi-Markov process approaches of chapter 4 are now also viable, simple, and usable descriptions of organic matter decomposition. I've shown how such complex systems can be described with just a single function, $v(\ln k)$, having a few parameters. The inversion of decay data to estimate the shape of the exit rate function is also a new procedure for parameter estimation of decay systems and offers a novel way to measure rate heterogeneity.

The more general abstraction of the decay system as a quasi-continuum network also serves as a loose connection to the diverse biological, physical and chemical processes underlying these systems. But from this network description, various dynamic properties of decay systems emerge. These include the description of mass loss as a continuous superposition of exponential decays, decay characterized by a lognormal distribution of rates, apparent instantaneous decay rates which decrease with time, common structures in observed kinetic heterogeneity, and turnover times and mean ages which depend on rate heterogeneity. In order to predict mass loss, any level of abstraction can be used. Often, it is most convenient to use the Laplace transform of an exit rate function or the higher level lognormal $p(k)$ to calculate $g(t)$. Models from these levels were applied in chapter 6 for example to implement the effects of climate on degradation. Suggestions for incorporating these concepts into large scale models are also presented at the end of chapter 6.

Decomposition on some scales appears simple. However, predicting kinetics based on 'first principles of decomposition mechanics' remains a challenge. The model is phenomenological, as it quantifies our qualitative understanding of soil organic matter as a heterogeneous continuum with mineralization derived from small-scale biological processes. It neglects however the non-linear processes which are present at small scale and over short times such as enzyme kinetics, microbial growth, sorption, etc. [7, 84, 147]. Rather than describing in detail the nonlinearities of some processes, the model more simply captures the heterogeneous kinetics of degradation to first order. The stochastic nature of decomposition systems and the collapse of complex system information onto a compact form, $v(\ln k)$, suggests that more detailed mechanistic approaches may be inefficient or even unnecessary for predicting decay.

Ecosystem ecology is naturally related to these decomposition networks. The effect of natural selection on decomposition rates discussed in chapter 5 is a striking and profound result. The work of this thesis regards the dynamics of organic matter respiration and is relevant to the interaction of the carbon cycle and climate. Yet it also yields information regarding the behavior and properties of networks associated with the newer field of systems ecology. Continued work in this area may reveal how such biological networks and their evolution has driven energy and mass flow on the earth over the past few billion years.

Appendix A

Residence Times, Ages, and Turnover of Mass in Linear Time-Invariant Systems

This chapter organizes and presents clearly the statistical theory of fluxes of matter in linear systems with time-invariant(LTI) dynamics,

$$\frac{dG(t)}{dt} = J(t) - F(t) \tag{A.1}$$

where G is the dynamic quantity of interest (here it is mass), $J(t)$ is a time varying input and the output is $F(t)$. The internal microscale dynamics of time-invariant systems do not change with time since all matter entering the system is always processed in the same manner. These systems are easily analyzed using *impulse responses.* Other statistical properties of matter flux are derived from the impulse response function.

A.1 Impulse Response Function

We assume here that an impulse of matter has arrived at a reservoir and that the governing dynamics of matter in the reservoir are known. $G(t)$ is the resulting dynamics of an impulse of size $G(0)$ to a system. The rescaled dynamics $g(t) = G(t)/G(0)$ is the *impulse response function.* The impulse response function is dimensionless and can be used to determine the dynamics of a *linear time invariant system* (A.1) receiving a non-steady input $J(t)$ by convolving the input with the impulse response function,

$$G(t) = \int_0^t J(\tau)g(t-\tau)d\tau \tag{A.2}$$

Since $g(t)$ completely describes the dynamics of the system we can extract additional useful information, such as the distribution of residence times that incoming matter will have in the reservoir, or age distributions of matter in the reservoir.

A.2 Residence Time Distributions

After an impulse $G(0)$ at time $t = 0$, matter continuously leaves the system. We can quantify this from the impulse response $g(t)$ by noting that an infinitesimal amount of matter $\delta G(t) = G(0)\delta g(t)$ exits the reservoir in the time interval $[t, t + \delta t]$ after an input. From the impulse response curve,

$$\delta g(t) = -\frac{dg}{dt}\delta t \tag{A.3}$$

and we say that $\delta G(t)$ had a residence time of t. The minus sign is required in (A.3) since $\delta g(t)$ is the *amount* of matter exiting the reservoir. The residence time t is weighted by mass slice $\delta G(t)$ which results in a distribution of residence times $\delta G(t) = G(0)\delta g(t)$. Normalizing $\delta G(t)$ by the total mass $G(0)$ gives the *residence time probability distribution function (pdf)*

$$\phi(t) = -\frac{dg}{dt} \tag{A.4}$$

Integrating $\phi(t)$ from $0 \leq \tau \leq t$, we find of course that $\int_0^t -\frac{dg}{d\tau}d\tau = -\int_1^g dg = 1 - g$ which is the matter lost from the reservoir at time t. Since $\phi(t)$ is a pdf, it integrates to 1 when $t = \infty$.

$$\int_0^\infty \phi(\tau)d\tau = 1 \tag{A.5}$$

We see now that the impulse response function is also related to the cumulative distribution function (cdf) of the residence times $\Phi(t) = \int_0^t \phi(\tau)d\tau$.

$$g(t) = 1 - \Phi(t) \tag{A.6}$$

The mean residence time is the first moment of $\phi(t)$

$$\langle\tau\rangle = \int_0^\infty \tau\phi(\tau)d\tau \tag{A.7}$$

A.3 Flux Out and Turnover Time

Reservoir dynamics are governed by the impulse response function (A.2). The *flux out of the system* $F(t)$ can be also calculated as continuous superposition of impulse response fluxes.

$$F(t) = \int_0^t -dg/d\tau J(\tau)d\tau \tag{A.8}$$

At steady state of course

$$\begin{align} F(t) &= \int_0^\infty J(\tau)(-dg/d\tau)d\tau \tag{A.9}\\ &= J \tag{A.10}\end{align}$$

The mass of the reservoir $G(t)$ (A.2) in steady state is

$$\begin{align} G &= J\int_0^\infty g(t)dt \tag{A.11}\\ &= JT \tag{A.12}\end{align}$$

The *turnover time* of the reservoir is defined as $T(t) = G(t)/F(t)$ and in steady state is $T(t) = T = G/J$. The turnover time gives a typical timescale for matter to travel through the system. In fact, *the turnover time is equal to the mean residence time in steady state.* This is a very intuitive statement, nonetheless I will show it mathematically.

First, calculate the mean residence time by integrating (A.7) over dg,

$$\begin{align} \langle\tau\rangle &= -\int_1^0 \tau(g)dg \tag{A.13}\\ &= \int_0^1 \tau(g)dg \tag{A.14}\\ &= A_g \tag{A.15}\end{align}$$

Where A_g is the area under the curve $\tau(g)$. Since $g(\tau)$ is essentially a cdf, the average value of τ is $A_g/(1-0) = A_g$. Therefore $A_g = \mathcal{T}$ the average value of $\tau(g)$ resulting in

$$\langle\tau\rangle = \mathcal{T} \tag{A.16}$$

The relation (A.16) relates a cdf and a pdf and is a property of all probability distributions. Similarly, the steady state turnover time relation (A.11) states

$$\begin{align} T &= \int_0^\infty g(\tau)d\tau \tag{A.17}\\ &= A_g, \tag{A.18}\end{align}$$

where T is also the area A_g under the curve τg. Since $A_g = \mathcal{T}$ it follows that in steady state

$$T = \mathcal{T} = \langle\tau\rangle. \tag{A.19}$$

Since the system is linear, *the mean residence time of the inflow is constant even if the input* $J(t)$

is non-steady. However the mass and flux out of the system may vary disproportionately with a dynamic input resulting in $T(t)$.

A.4 Age Distribution

The distribution of ages in a reservoir is also an important property and is not necessarily the same as the residence time distribution. An input $J(\tau)d\tau$ will have a age $a = t - \tau$, where t is the current time and τ is the time of the input. The amount of matter in the reservoir with age a is $J(\tau)d\tau g(t-\tau)$. Therefore the distribution of ages of matter in the reservoir is $J(t-a)g(a)$. Normalizing by the total mass gives the pdf of age distribution

$$\alpha(a,t) = J(t-a)g(a)/G(t) \tag{A.20}$$

In steady state the age pdf is

$$\alpha(a) = J/Gg(a) \tag{A.21}$$
$$\alpha(a) = g(a)/T. \tag{A.22}$$

and the mean age is

$$\langle a \rangle = \int_0^\infty ag(a)/Tda. \tag{A.23}$$

Case of equal residence times and ages

In some situations, it is desirable to substitute the residence time distribution by the age distribution or vice versa. However this is not always a valid substitution. In order for $\phi(\tau) = \alpha(a)$,

$$g(a)/T = -dg(a)/da \tag{A.24}$$
$$dg(a)/da = -g(a)/T \tag{A.25}$$
$$g(a) = e^{-t/T}, \tag{A.26}$$

the impulse response must be exponential with a time constant T, the turnover time. The resulting mean age and mean residence time is therefore of course

$$\langle a \rangle = \int_0^\infty t/Te^{-t/T}dt \tag{A.27}$$
$$= T \tag{A.28}$$

A.5 Case of Random Rates

When the impulse response function is completely monotonic (section 2.7.1), it is possible to describe $g(t)$ as a superposition of a distribution of first order decay rates.

$$g(t) = \int_0^\infty p(k)e^{-kt}dk \tag{A.29}$$

The underlying physics of the system may or may not be related to the disordered decay (A.29). However, the underlying physics of the system may be more related to the Laplace transform of a more general exit rate function $r(k)$. An exit rate function derives from any non-degenerate distribution of residence times, and thus $g(t)$ can take the shape of any monotonically decreasing function. Reservoir mass, outflux, residence time distribution and age distribution can all be calculated from the distribution of reaction rates $p(k)$ or exit rate function $r(k)$.

All of the following analysis applies to both $p(k)$ and $r(k)$. $r(k)$ may be substituted for $p(k)$ in any of the following equations.

A.5.1 Mass with distributed rates

Substituting (A.29) for $g(t)$ in the convolution (A.2) gives $G(t)$

$$G(t) = \int_t^0 J(t-\tau)\int_0^\infty -p(k)e^{-k\tau}dkd\tau \tag{A.30}$$

In the steady state, we can find $G(t)$ by integrating through time.

$$G = J\int_0^\infty \int_\infty^0 -p(k)e^{-k\tau}d\tau dk \tag{A.31}$$

$$= J\int_0^\infty p(k)k^{-1}dk \tag{A.32}$$

A.5.2 Outfluxes and turnover times associated with distributed rates

The outflux $F(t)$ is found by inserting (A.29) into (A.8)

$$F(t) = \int_0^t \left\{\int_0^\infty -kp(k)e^{-k\tau}dk\right\} J(\tau)d\tau \tag{A.33}$$

which simplifies in steady state to J. The turnover time is of course $G(t)/F(t)$ and in steady state is

$$T = G/F \tag{A.34}$$

$$= \int_0^\infty p(k)k^{-1}dk \tag{A.35}$$

where we have divided mass (A.32) by the flux J.

A.5.3 Residence times with distributed rates

Substituting (A.29) into the residence time distribution (A.4)

$$\phi(\tau) = \int_0^\infty kp(k)e^{-k\tau}dk \tag{A.36}$$

$\phi(0)$ gives the mean rate $\langle k \rangle$. The mean residence time however is not $\langle k \rangle^{-1}$. It can be calculated by finding the first moment of $\phi(\tau)$ (A.36)

$$\begin{aligned} \langle \tau \rangle &= \int_0^\infty \int_0^\infty k\tau e^{-k\tau} d\tau p(k)dk & \text{(A.37)} \\ &= \int_0^\infty k^{-1}p(k)dk & \text{(A.38)} \end{aligned}$$

Which is simply the mean of the relaxation times $t_r = 1/k$ which is easily seen by substituting the pdf of t_r for $p(k)$. Again we see this is the same as the turnover time in steady state (A.35).

A.5.4 Age distribution associated with distributed rates

We can also write the age distribution as a function of $p(k)$.

$$\alpha(a,t) = I(t-a)\int_0^\infty p(k)e^{-ka}dk/G(t) \tag{A.39}$$

This readily simplifies in steady state to $g(a)/T$ as discussed above.

$$\alpha(a) = \int_0^\infty p(k)e^{-ka}dkT^{-1} \tag{A.40}$$

The mean age is found from the first moment of a

$$\langle a \rangle = T^{-1}\int_0^\infty k^{-2}p(k)dk \tag{A.41}$$

A.5.5 Linear time invariant properties of a lognormal distribution of k

If $p(k)$ is a lognormal distribution $\Lambda(k;\mu,\sigma)$, then dynamic reservoir masses, fluxes, residence times and ages may be found by substituting the lognormal for $p(k)$. The dynamic relations must be calculated numerically. The following steady state quantities however can be found analytically.

The steady state reservoir mass is

$$G = Je^{-\mu+\sigma^2/2} \tag{A.42}$$

Steady state turnover time and mean residence time is therefore

$$T = \langle \tau \rangle = e^{-\mu + \sigma^2/2} \tag{A.43}$$

Steady state mean age is

$$\begin{aligned} \langle a \rangle &= 1/T e^{-2\mu + 2\sigma^2} && \text{(A.44)} \\ &= e^{-\mu + 3/2\sigma^2} && \text{(A.45)} \\ &= T e^{\sigma^2} && \text{(A.46)} \end{aligned}$$

Again as discussed section A.4, the mean age is only the turnover time in the event that the impulse response is a single exponential, in other words $p(k) = \delta(t - \mu)$, $\sigma = 0$.

A.6 Summary

The impulse response function, residence time distribution, age distribution and turnover time can each be described and thought about in various ways. Various interpretations of each property are discussed here as well as a chart to summarize the results from this chapter.

$g(t)$ can be thought of as the decay from an impulse. It is the impulse response function. It can also be thought of as related to the cdf of the residence time distribution.

$\phi(\tau)$ is a residence time distribution. It is the distribution of lifetimes associated with the input stream. It is also the age distribution of the output stream. Since it represents the flux out of the system it is also the negative slope of $g(t)$.

$\alpha(a, t)$ is the age distribution of matter in the reservoir at time t. In steady state it is simply $g(t)$ normalized by the timescale T.

The turnover time $T(t)$ is the mass in the reservoir divided by the flux out. In steady state this is equal to the mean residence time and for the case of an exponential $g(t)$ it is also the mean age.

The following table summarizes the calculations of this chapter as functions of J, $g(t)$ and $p(k)$.

A.7 Predicting Soil Carbon Stores and Age in Field Experiments

It is difficult to find a long term litter decomposition study that also included a local measurement of soil-C storage, estimated C input and measured mean age of soil C. Jenkinson performed a similar experiment in a field in Rothamstead, England during the 1970's. Jenkinson monitored the decay of C^{14} labeled ryegrass in a field in for ten years. He fit a 2 component decay and predicted the mean age and turnover time from the model. The Broadbalk fields where the ryegrass decayed had been used to grow winter wheat since 1881. A Broadbalk soil sample from 1881 had a carbon

Property	Dynamic	Steady	Random Rate*	Lognormal*
$G(t)$	$\int_0^t J(\tau)g(t-\tau)d\tau$	$J\int_0^\infty g(t)dt$	$J\int_0^\infty p(k)k^{-1}dk$	$Je^{-\mu+\sigma^2/2}$
$F(t)$	$\int_0^t dg/d\tau J(\tau)d\tau$	J	J	J
$T(t)$	$G(t)/F(t)$	$\int_0^\infty g(t)dt$	$\int_0^\infty p(k)k^{-1}dk$	$e^{-\mu+\sigma^2/2}$
$\phi(\tau)$	$-dg(\tau)/dt$	$-dg(\tau)/dt$	$\int_0^\infty kp(k)e^{-k\tau}dk$	$\int_0^\infty k\Lambda(k)e^{-k\tau}dk$
$\langle\tau\rangle$	$\int_0^1 \tau(g)dg$	$\int_0^1 \tau(g)dg$	$\int_0^\infty k^{-1}p(k)dk$	$e^{-\mu+\sigma^2/2}$
$\alpha(a,t)$	$I(t-a)g(a)/G(t)$	$g(a)/T$	$\int_0^\infty p(k)e^{-ka}dk/T$	$\int_0^\infty \Lambda(k)e^{-ka}dk/T$
$\langle a\rangle$	$\int_0^\infty a\alpha(a,t)da$	$\int_0^\infty ag(a)/Tda$	$\int_0^\infty k^{-2}p(k)dk/T$	Te^{σ^2}

Table A.1: Turnover time, Flux, Mass, Residence Time and Age distributions of linear time invariant systems. *indicates steady state calculations. For non-steady calculations see the text.

content of 26 t/ha and a also mean age of about 1450 yr using δC^{14}. An 1881 soil was analyzed because it was before the bomb-carbon spike around 1955. Decomposition from the ryegrass decay experiment were used with pre and post measurements of soil δC^{14} to estimate an input of organic matter of about 1.2 t/(ha-yr) during the continuous winter wheat experiment in the 1950s. This resulted in a turnover time of about 22 years for unfertilized plots.

Comparing measured turnover times to the two component decay model resulted in a severe underfit of the estimated soil-C turnover time and mean age for multiple reasons. The main reason is that the two component model did not have a slow pool. However the LIDET study has shown that substrate type is very important and can strongly control decay dynamics. The vegetation growth on the Rothamstead plot from 1843-1881 was winter wheat with various fertilizers. Broadbalk is believed to have been farmland for centuries before 1843 [149]. The decay of ryegrass therefore cannot necessarily predict the C-store and mean soil-C age in 1881 of random farm crops grown for a few centuries and remains of previous vegetation. The estimate of input of winter wheat using ryegrass decomposition parameters and dividing by an 1881 soil-C store to get turnover time is also not necessarily related to the turnover time predicted by the 2-component model of ryegrass decomposition. The winter wheat may have markedly different values of μ and σ and carbon influx was determined by bomb carbon isotope and the ryegrass data which likely results in error in turnover time estimation.

Nevertheless, analyzing this data is a useful exercise to demonstrate how the random rate model predicts long term soil-C dynamic properties. Fitting the lognormal to the ryegrass decay results in a mu of 1.07 and sigma of 3.2. This results in a turnover of 56 yr and 1.6 million year mean age. The age seems to have a super heavy tail (with only 1ppm remaining at the mean age). The mean age must take a while to reach steady state so a numerical investigation of the age distribution as a function of time should be checked to see how long it takes for a mean age of 1330 yr to be reached. Farming started in England around 6000 years ago, http://www.fsmitha.com/h1/ch07.htm The LIDET dataset contained some spring wheat replanted at its home location of Kellogg forest which has a mean annual temperature equal to Rothamstead and a precipitation of about 920 mm/yr vs Rothamstead's 704mm/yr. The spring wheat at Kellogg had $\mu = -.76$ and $\sigma = .75$ corresponding to a turnover time of 2.8 yr and mean age of 4.9 year. Winter wheat is more refractory than spring wheat.

Appendix B

Lognormally Distributed Rates and their Approximation

The consistency of the inversion with the lognormal distribution leads us to model the reaction rates in litter as lognormally distributed. However, in order to predict mass dynamics with this model, the Laplace transform of the lognormal needs to be calculated. Because it has no analytical solution, I consider various approximations to the Laplace transform of the lognormal in the following two chapters.

As in section 3.5, $g(t)$ is predicted by substituting the lognormal distribution $\Lambda(k, \mu, \sigma)$ with mean log rates μ and variance of log rates σ^2 into (3.1)

$$g(t) = \int_0^\infty \Lambda(k, \mu, \sigma) e^{-kt} dk, \tag{B.1}$$

The decay model (B.1) can be reduced via eliminating parameter μ by rescaling time by the median reaction rate $\tau = k^* t$ and $\kappa = k/k^*$ where $k^* = e^\mu$

$$g_\sigma(\tau) = \int_0^\infty \Lambda(\kappa, 0, \sigma) e^{-\kappa\tau} d\kappa \tag{B.2}$$

Typically, equation (B.1) or (B.2) must be evaluated numerically in $\log k$ space in order to determine $g(t)$ except for certain cases of small or large σ. Approximations for these cases are discussed in section B.2 and in chapter C.

B.1 Decay Model Fits LIDET Data

The two parameter decay model was fit to a subset of 232 LIDET data sets which did not exhibit mass accumulation or negligible mass loss over the length of the study and had at least 6 data points. Figure B-1 depicts rescaled decay data, where I've rescaled time by $\tau = e^\mu t$ and mass is

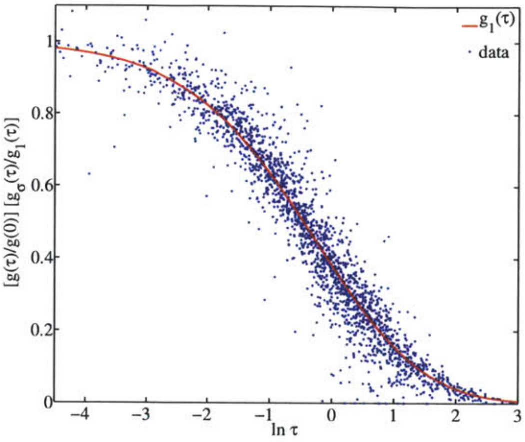

Figure B-1: 232 rescaled LIDET data sets. Dots are average mass fraction remaining from litter bags of the same substrate removed at the same time and location. Red line is decay resulting from the $\Lambda(k;0,1)$ distribution.

rescaled by the mass remaining from the unit lognormal distribution $\Lambda(k;0,1)$. To rescale mass, the mass fraction is multiplied by the σ-dependent quantity $g_\sigma(\tau)/g_1(\tau)$. Although all of the data can be rescaled to the same curve, the decay curves actually can exhibit markedly different behaviors. Some datasets show exponential decay where $\sigma \ll 1$, some reveal decay with $\log t$ where $\sigma \gg 1$, and the remaining have moderate values of σ and are somewhere between. Figure B-2 shows examples of those three cases plotted on a linear log plot. The $\sigma \ll 1$ case is exponential, while the case of $\sigma \approx 2$ already looks like $log(\tau)$ decay over the range of the data.

B.2 Asymptotic Approximations of Lognormal Decay

$g(t) \rightarrow$ exponential decay as $\sigma \rightarrow 0$ simply because there is only one rate $k^* = e^\mu$ in the reactive continuum. Simply replace $p(k)$ with a delta function in equation (B.1).

$$g(t) = \int_0^\infty \delta(k - k^*) e^{-kt} dk \tag{B.3}$$

$$g(t) = e^{-k^* t} \tag{B.4}$$

$g(t) \rightarrow$ logarithmic decay with time as $\sigma \rightarrow \infty$ because only a small portion of the continuum is actively decaying at any moment in time and the resulting Laplace transform is approximately linear with $\log t$ for most of the degradation of g. This is most easily shown by first changing variables from k to $\lambda = \log k$

$$g(t) = \int_{-\infty}^{\infty} N(\lambda; \sigma, \mu) e^{-e^\lambda t} d\lambda \tag{B.5}$$

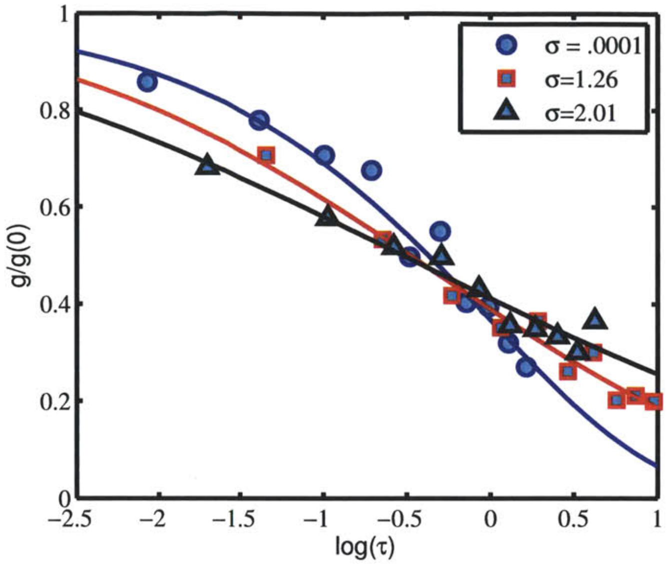

Figure B-2: Mass fraction remaining vs $\log \tau$ are plotted for three LIDET data sets with different values of σ. Solid lines are predictions of decay from a lognormal distribution of rates for each value of σ. Three types of decay are present: exponential $\sigma = .0001$, logarithmic with time $\sigma = 2$, and a moderate decay $\sigma = 1.26$

As $\sigma \to \infty$, the exponential decay $e^{-e^{\lambda}t}$ in (B.5) approaches a Heaviside step function.

$$e^{-e^{\lambda}t} \approx H(log(t) - \lambda) \tag{B.6}$$

In this manner, the Laplace transform acts as a filter. This is the motivation for the constraint used in section 4.10.7. To see this perhaps more clearly: at any given time t, the portion of the continuum with rate $k = 1/t$ has decayed to a fraction e^{-1} of its original value. Matter with rates $k \gg 1/t$ have decayed almost entirely, and organic matter with rates $k \ll 1/t$ have hardly changed from their original values. The error in this Heaviside approximation is

$$\delta g_e = \int_{-\log t}^{\infty} \rho(\lambda)(1 - e^{-e^{\lambda}t})d\lambda - \int_{-\infty}^{-\log t} \rho(\lambda)(e^{-e^{\lambda}t})d\lambda. \tag{B.7}$$

Where it is easily seen that the integrand decays to zero quickly when integrating away from $\lambda = \log t$. If $\sigma \gg 1$ then the the integrand in (B.7) is non-zero over such a relatively small range of λ compared to σ that $g(t) \gg \delta g_e(t)$.

Using the Heaviside approximation, the decay in the case of $\sigma \gg 1$ can also be calculated analytically.

$$g(t) = \int_{-\infty}^{\infty} \rho(\lambda) H(\log t - \lambda) d\lambda \tag{B.8}$$

$$g(t) = \int_{-\infty}^{\log t} \rho(\lambda) d\lambda \tag{B.9}$$

$$g(t) = 1 - P(\log t) \tag{B.10}$$

where P is the cumulative distribution function of pdf ρ. So, generally any very wide distribution $\rho(\lambda)$ will have a Laplace transform related to its cdf. In the case of the normal distribution $\rho(\lambda)$, substituting $N(\lambda; \mu, \sigma)$ into (B.8) we find

$$g(t) = \frac{1}{2} - \frac{1}{2}\mathrm{erf}\left[\frac{\log t - \mu}{\sigma\sqrt{2}}\right] \tag{B.11}$$

Because the error function is sigmoidal, much of the mass loss lies along a rather linear portion and appears to decay like $\log t$ as will be seen in the next chapter. But even when width of the distribution is moderate such as $\sigma = 2$, $\log t$ seems to describe the measured decay data seen in figure B-2.

A caveat for the cdf approximation is that it eventually fails once $g \sim \delta g_e$. This will happen in the left tail of a distribution. Since tails must converge at least as fast as a power law with exponent < -1, no matter how large σ is, there will be a point where the small fraction $g(t) \sim \delta g_e$. Failure of this approach can also happen if $\rho(\log k)$ has a significant fraction of the continuum within a narrow range of rates. For example, if most of the data is contained in one delta function and a small fraction is out at ∞. A numerical calculation of $g(t)$ is necessary if far tail resolution is required or ρk has a wide variance but most of its continuum in a narrow range of rates.

B.3 Analytical Solutions to the Laplace Transform of the Lognormal Distribution

Laplace transform of the lognormal distribution cannot be represented by its moment generating function since the series does not converge [150]. Beck and Cohen [151] give a series solution but this is only a modified moment generating function and is therefore only valid for $\sigma \ll 1$ when terms are truncated. A general series solution to the Laplace transform of the lognormal exists but does not provide much physical insight and many terms are required for precision [152]. A series solution also exists for the characteristic function [150] but is not valid on the imaginary axis. Another asymptotic expansion is discussed by Holgate [153] (see section C.4) which is valid in the complex plane however it only well approximates the Laplace transform as $\sigma \to 0$ and over a finite range of time. Holgate also found a series solution to the characteristic function based on boundary analytic theory and claims that a similar calculation of the Laplace transform is possible. The Laplace transform of the lognormal is also related to the problem of software reliability [154]. Montroll [155] claims the Laplace transform of the lognormal distribution is approximately a stretched exponential for $\sigma \gg 1$ however upon investigation the stretched exponential only appears to be a good approximation for about 10% to 20% of the decay. The error function (B.11) is a much better approximation for $g(t)$ when $\sigma \gg 1$. Since the Laplace transform of the lognormal is inconvenient, the gamma distribution is often used as an approximation to the lognormal [156]. The Laplace transform of the lognormal can also be rewritten in terms of a zeroth Poisson-lognormal distribution [157], and was tabulated by Grundy in 1951. [158]. However since Laplace transforms of long-tailed distributions are easy on log-scales [159] we calculate $g(t)$ using numerical integration.

B.4 Loguniform

As discussed in section B.3, calculating g(t) from the lognormal reactivity distribution is mathematically inconvenient and requires either many terms of a complex series solution or numerical integration. It may be useful however to use an approximate reactivity distribution which produces an analytical $g(t)$. The lognormal distribution (2.16) can be rewritten as

$$\rho(k) = \frac{1}{\sqrt{2\pi}\sigma k} e^{\mu^2/2\sigma^2} k^{(2\mu - \ln k)/2\sigma^2} \tag{B.12}$$

From (B.12) it is easy to see that when $\sigma \gg 1$, $\rho(k) \sim 1/k$ for a wide range of k. Over the range $\mu - \sigma \ll \log\sqrt{k} \ll \mu + \sigma$, the lognormal is approximated by the loguniform distribution [120] $\sqcap(k)$, which has also been shown to describe the organic matter decay continuum [50],

$$\sqcap(k) = \begin{cases} (\ln k_{max}/k_{min})^{-1} k^{-1} & : k_{min} < k < k_{max} \\ 0 & : k < k_{min}, k > k_{max} \end{cases} \tag{B.13}$$

However, the Laplace transform of the lognormal requires an approximation across all k and section B.4.2 shows that the approximating $g(t)$ by the loguniform gets worse as $\sigma \gg 1$. Nevertheless, this approximation is equivalent to approximating the normal distribution of $\log k$ to zeroth order: a constant over a finite range of $\log k$. A constant distribution in $\log k$ space is a $1/k$ distribution in linear space in accordance with change of variables equation $p(k)dk = \rho(\ln k) d\ln k$.

B.4.1 Relations between lognormal and loguniform parameters

The two parameters k_{max} and k_{min} are set by the two parameters μ and σ of the lognormal distribution. A symmetry relation gives

$$\ln k_{max} + \ln k_{min} = 2\mu \tag{B.14}$$

which means that the lognormal and loguniform distribution are centered same mean in log-space and same median in linear and log-space.

Empirically, we also find that

$$\ln k_{max} - \ln k_{min} = 3.1\sigma. \tag{B.15}$$

Relation (B.15) is based on comparing a best fit of k_{max} and k_{min} to the best fit σ over the decays in the data set as well as synthetic data generated from a lognormal distribution. This scaling between σ and $\ln(k_{max}/k_{min})$ can also be found analytically by minimizing the error ε between the lognormal and loguniform in log space.

$$\begin{aligned}
\varepsilon &= 2\int_{\infty}^{0} (\mathcal{N}(\lambda;0,\sigma) - \Pi(\lambda;a))^2 d\lambda && \text{(B.16)}\\
&= 2\int_{-\infty}^{0} \mathcal{N}(\lambda)d\lambda + \int_{-a}^{0} (4a^2)^{-1} - \mathcal{N}(\lambda)/a d\lambda && \text{(B.17)}\\
&= (2\sqrt{\pi}\sigma)^{-1} + (2a)^{-1} - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}\right) a^{-1} && \text{(B.18)}
\end{aligned}$$

where $2a = \ln(k_{max}/k_{min})$ and $\Pi(\lambda)$ is the loguniform distribution in log-rate space. It is calculated from $\sqcap(k)$ by the Jacobian $\frac{d\sqcap(k)}{dk} = \frac{\Pi(\lambda)}{d\lambda}$. To find the a which minimizes ε, set $\frac{d\varepsilon}{da} = 0$

$$\begin{aligned}
\frac{d\varepsilon}{da} &= e^{-a^2/2\sigma^2}\sqrt{2}(\sqrt{\pi}\sigma a)^{-1} - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}\right) a^{-2} - (2a^2)^{-1} && \text{(B.19)}\\
0 &= 1/a^2\left(e^{-a^2/2\sigma^2}\frac{\sqrt{2}a}{\sqrt{\pi}\sigma} - \operatorname{erf}\left(\frac{a}{\sqrt{2}\sigma}\right) - 1/2\right). && \text{(B.20)}
\end{aligned}$$

After making a^2 the common denominator of all terms in (B.19), the parameter a always appears in the numerator with $1/\sigma$. Maple's numerical root finder `fsolve` is used to solve transcendental equation (B.19) for a/σ

$$\begin{aligned}
a/\sigma &= 1.538 && \text{(B.21)}\\
\ln(k_{max}/k_{min})/\sigma &= 3.08 && \text{(B.22)}
\end{aligned}$$

which is in agreement with the empirical finding (B.15). The numerical calculations presented in the next section involving both the lognormal and loguniform distributions use (B.15) and (B.14) to relate k_{max} and k_{min} to σ.

B.4.2 Comparing the Laplace transforms of the lognormal and loguniform distributions

The Laplace transform of the loguniform best approximates the Laplace transform of the lognormal when $\sigma \ll 1$. This is because when $\sigma \to 0$, both $\Lambda(k)$ and $\sqcap(k)$ approach a delta function at $k = e^{\mu}$ and decay in both cases is exponential. Differences between the Laplace transforms of the two distributions are apparent for larger σ since the timescales of the error are widely spread, resulting in a time dependent error.

Decay Curves

When $\sigma \to \infty$ the Laplace transforms of the two distributions are most dissimilar, exposing the full extent of the error in the approximation. Since $\sigma \to \infty$ we can approximate the the error in the Laplace transform of the distributions by using the cdfs of $\rho(\lambda)$ as discussed in section B.2.

$$\begin{aligned} E(t) &= \mathscr{L}(\Lambda(k) - \sqcap(k)) && \text{(B.23)} \\ &= (1 + \operatorname{erf}\{(-\ln t - \mu)/\sqrt{2}\sigma\})/2 - \begin{cases} 1 & : t^{-1} > k_{max} \\ \ln(tk_{min})/\ln(k_{max}/k_{min}) & : k_{min} < t^{-1} < k_{max} \\ 0 & : t^{-1} < k_{min} \end{cases} && \text{(B.24)} \end{aligned}$$

However $\mathscr{L}(\sqcap(k))$ has an analytical solution and has also been used to describe organic matter decay [50]. Rothman and Forney showed that a steady reaction diffusion equation in an absorbing porous medium results in a loguniform distribution of reaction rates $\sqcap(k)$ and its Laplace transform is

$$\mathscr{L}(\sqcap(k)) = (\ln(k_{max}/k_{min}))^{-1}(\mathrm{E}_1(k_{min}t) - \mathrm{E}_1(k_{max}t)) \tag{B.25}$$

Similarly, for $\log k_{max}/k_{min} \gg 1$, $\mathscr{L}(\sqcap(k))$ decreases like $\log t$ between k_{max}^{-1} and k_{min}^{-1}.

The error can then be written more precisely as

$$E(t) = \mathscr{L}(\Lambda(k)) - (\ln(k_{max}/k_{min}))^{-1}(\mathrm{E}_1(k_{min}t) - \mathrm{E}_1(k_{max}t)), \tag{B.26}$$

and for $\sigma \gg 1$

$$E(t) = (1 + \operatorname{erf}[(-\ln t - \mu)/\sqrt{2}\sigma])/2 - (\ln(k_{max}/k_{min}))^{-1}(\mathrm{E}_1(k_{min}t) - \mathrm{E}_1(k_{max}t)). \tag{B.27}$$

To measure how well decay from a loguniform distribution approximates the decay from a lognormal

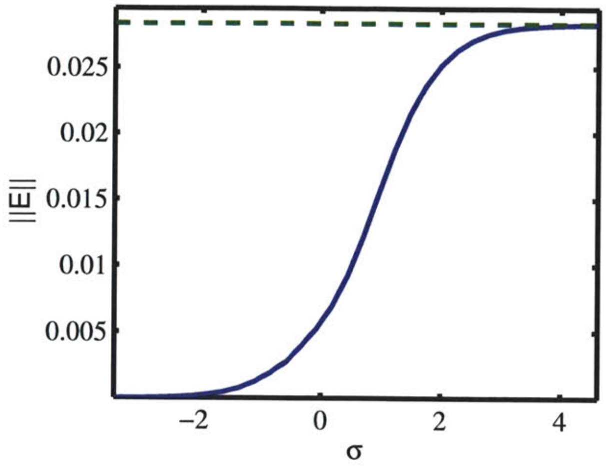

Figure B-3: $\|E\|$ numerically integrated over 98% of the decay. $\sigma \gg 1$ has larger error. The dashed line is calculated by substituting (B.27) for $E(t)$ in RMSE equation (B.28) and is the asymptote as $\sigma \to \infty$.

distribution, we calculate the root mean square error (RSME) in log time over 98% of the decay.

$$\|E\| = \sqrt{\int_{\log T_1}^{\log T_2} E(t)^2 d\log t / \log{(T_2/T_1)}} \tag{B.28}$$

where $g(T1)/g(0) = .99$ and $g(T2)/g(0) = .01$. Using the error formulation (B.26) in (B.28), $\|E\|$ is numerically integrated and plotted in figure B-3.

As σ increases, the discrepancy between a lognormal and a loguniform distribution becomes more apparent and reaches an asymptotic maximum as $\sigma \to \infty$. The asymptote is be calculated from the cdf relation (B.27) and is $\|E\| = .028$. Figure B-4 shows both $\mathscr{L}(L(k))$ and $\mathscr{L}(\sqcap(k))$ (B.25) vs $\ln t$ when $\sigma \gg 1$. In this worst case comparison, the log time RMSE is .028 with a maximum error of $(g_\Lambda - g_\sqcap)/g(0) = .062$.

Another difference associated with using the loguniform distribution is that the loguniform predicts a relatively stable mass fraction remains long after the loguniform predicts a relatively complete decay. At $t = 1/k_{min}$

$$\begin{aligned}
E(1/k_{min}) &\approx (1 + \operatorname{erf}[(\ln k_{min} - \mu)/\sqrt{2}\sigma])/2 && \text{(B.29)}\\
&= (1 + \operatorname{erf}[(\mu - a - \mu)/\sqrt{2}\sigma])/2 && \text{(B.30)}\\
&= 1/2 + \operatorname{erf}(-1.54/\sqrt{2})/2 && \text{(B.31)}\\
&= .062 && \text{(B.32)}
\end{aligned}$$

which will remain for much longer than $t = 1/k_{min}$.

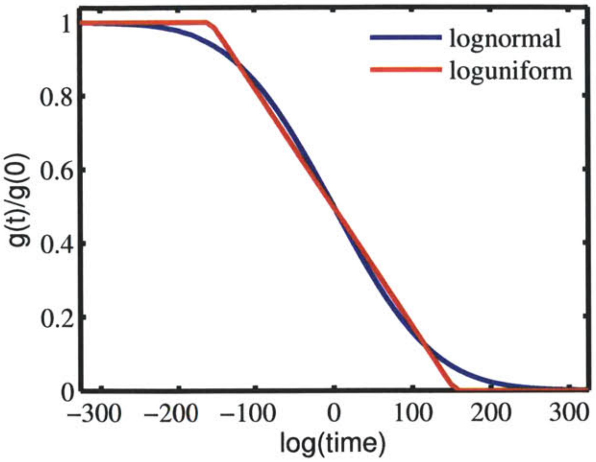

Figure B-4: $L(t)$ vs $\ln t$, blue. $\mathscr{L}(\sqcap(k))$ vs. $\ln t$, green. Both for $\sigma = 100$. Shown here is the largest discrepancy between the decay of a lognormal distribution and a loguniform distribution. Integrating the error between these two curves gives approximately the asymptote for $\|E\|$ in figure B-3. Also note that cdf $\Pi(t)$ well approximates $\mathscr{L}(\sqcap(k))$.

Turnover Times

A departure at large σ is also found when comparing lognormal and loguniform turnover times as discussed in chapter A. The steady state turnover time of a loguniform distribution is

$$T = \ln^{-1} k_{max}/k_{min}(1/k_{min} - 1/k_{max}). \tag{B.33}$$

The turnover time of the lognormal is shown in equation (2.5). Using equation (B.21) to relate σ and μ to k_{max} and k_{min} figure B-5 shows that turnover times from the two distributions are similar only when σ is small. When sigma is large, the turnover time is rapidly shifted out in the tail, while the log turnover time from the loguniform scales approximately linearly with σ and approaches $1/k_{min}$.

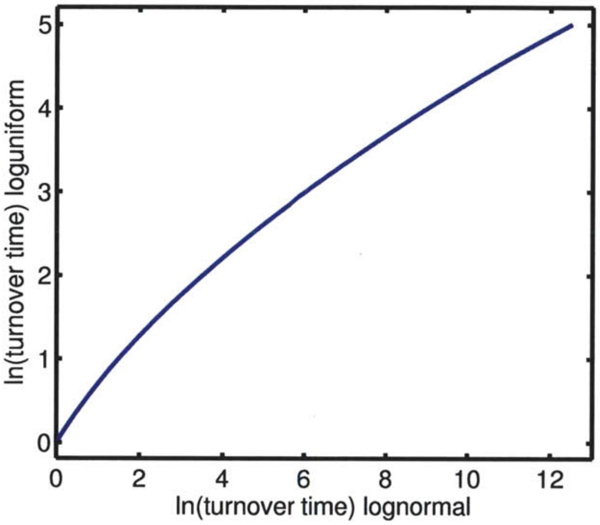

Figure B-5: ln turnover time of the lognormal vs ln turnover of the loguniform for $0 < \sigma < 5$. Time and rates have been rescaled by e^{μ} resulting in a nondimensional turnover time as a function of σ. Turnover times increase as σ increases from 0.

Appendix C

Approximation of the Laplace Transform of the Lognormal: Moment Closure

Abstract

Moment closure techniques have been used to approximate the Laplace transform of a distribution of decay rates by a set of differential equations [16,10]. The modified moment closure technique of Bruggeman [160,161] holds well for lognormally distributed rates with standard deviation $\sigma < 1$, however fails for larger σ even when more terms and moments are included. This moment closure technique provides an analytical solution which holds for larger σ than any previously published low order approximation to the Laplace transform of the lognormal distribution. More carefully closing the set of 3 ODEs results in an accurate prediction of decay up to $\sigma < 1.6$ but does not have an analytical solution. A set of 4 ODEs with carefully closed moments yields accurate numeric estimation of decay from lognormally distributed rates with $\sigma < 3$ but has no analytical solution.

C.1 Motivations

Moment closure techniques have been applied to lognormal rate distributions [10,161] and are of interest here because the Laplace transform of the lognormal has no closed form solution. The moment closure technique is of further interest because the carbon cycle models used in coupled climate-carbon cycle simulations often are implemented as differential equations.

C.2 Standard Moment Closure

Decay from a continuous distribution $p(k)$ of reaction rates k can be approximated by a set of differential equations which describe how moments change in time. This technique is called moment expansion or moment closure [10,160,162,163].

C.2.1 Set of raw moment equations

The mass remaining $g(t)$, mean rate $m(t)$ and raw moments $M_i(t)$ are

$$\begin{aligned} g(t)/g(0) &= \int_0^\infty p(k,t)dk \\ m(t) &= \int_0^\infty kp(k,t)dk \\ M_2(t) &= \int_0^\infty k^2 p(k,t)dk \\ M_3(t) &= \int_0^\infty k^3 p(k,t)dk \\ M_4(t) &= ... \end{aligned} \tag{C.1}$$

where $p(k,t) = p(k,0)e^{-kt}/g(t)$. To write down the differential equations governing the change in mass and moments we take the derivative of eq (C.2). To make this easier, multiply both sides of the moment equations by $g(t)$ before taking the derivative.

$$\begin{aligned} \frac{d}{dt}g(t)/g(0) &= -\int_0^\infty kp(k,0)exp(-kt)dk \\ \frac{d}{dt}g(t)m(t) &= -\int_0^\infty k^2 p(k,0)exp(-kt)dk \\ \frac{d}{dt}g(t)M_2(t) &= -\int_0^\infty k^3 p(k,0)exp(-kt)dk \\ \frac{d}{dt}g(t)M_3(t) &= -\int_0^\infty k^4 p(k,0)exp(-kt)dk \\ \frac{d}{dt}g(t)M_4(t) &= ... \end{aligned} \tag{C.2}$$

And by definition of the mean and raw moments

$$\begin{aligned} \frac{d}{dt}g/g(0) &= -gm \\ \frac{d}{dt}gm &= -gM_2 \\ \frac{d}{dt}gM_2 &= -gM_3 \\ \frac{d}{dt}gM_3 &= -gM_4 \\ \frac{d}{dt}gM_n &= -gM_{n+1} \end{aligned} \tag{C.3}$$

resulting in a simple set of ODEs to describe decay.

C.2.2 Set of central moment equations

We can similarly take time derivatives of the moments about the mean. Using the product rule to find $\frac{d}{dt}m(t)$

$$\frac{d}{dt}g(t)m(t) = g(t)\frac{d}{dt}m(t) + m(t)\frac{d}{dt}g(t) \tag{C.4}$$

and substituting (C.3) into (C.4)

$$\begin{aligned} \frac{d}{dt}m(t) &= m(t)^2 - M_2(t) && \text{(C.5)} \\ &= -v(t) && \text{(C.6)} \end{aligned}$$

where $v(t)$ is the variance about the mean. Similarly, the derivatives of the moments about the mean (central moments) m_i are

$$\frac{d}{dt}m_i(t) = -m_{i+1}(t) \tag{C.7}$$

resulting in the following simple set of equations for the system

$$\begin{aligned} \frac{d}{dt}g/g(0) &= -mg \\ \frac{d}{dt}m &= -v \\ \frac{d}{dt}v &= -m_3 \\ \frac{d}{dt}m_3 &= -m_4 \\ \frac{d}{dt}m_4 &= \dots \end{aligned} \tag{C.8}$$

C.2.3 Closing the moments

Moment closure can be attempted using many ways. For lognormally distributed $p(k)$, various moment closure methods were implemented. Unfortunately all of these methods fail for roughly $\sigma > 0.1$ as the solution becomes unstable and explodes at a finite crossover time with much mass remaining.

Closing with normal distribution

A standard way is to use the moments of a normal distribution to close the system. For example if in equations (C.8), I was using a set of 4 ODEs which track up to the 3rd moment and set the fourth moment m_4 to be a function of the variance, $m_4 = 3v$. This is obviously a problem since the distribution evolves from the initial distribution with time and is even more of a problem when the initial distribution is not normal. Furthermore if the number of ODEs is n, then m_{n+1} can be calculated from either $v, m_4, m_6, \ldots$ none of which would agree with one another. This method fails quickly.

Closing with the initial distribution

Calculating the next highest moment from the initial distribution $p(k,0)$ is also prone to error for the same reason: the distribution evolves with time.

For example, with p(k)=lognormal distribution, I used $m(t)$ and $v(t)$ to calculate the parameters of the lognormal distribution with matching mean and variance. These parameters were then used to calculate the higher moment required for closure. This becomes unstable quickly for $\sigma > .1$ as more terms are included.

Closing with an approximate distribution

This method was used to account for the change in skew in the closing distribution. I attempted to account for the change of skew with a three parameter lognormal distribution. $m(t), v(t), m_3(t)$ were used to calculate the parameters of the 3 parameter lognormal distribution from chapter 3 of [164]. The third parameter a simply shifts k to $k-a$. The parameters were then used to calculate m_4, however the solution becomes unstable for $\sigma > .1$.

Closing with a-priori knowledge of any moment

This will work if the dynamics of a moment $m_n(t)$ are exactly known. Any moment can be found by numerically calculating the moment $m_n(t)$ exactly at a given sigma, storing the result and then using it to solve the ODE. I didn't investigate this but perhaps a moment $m_n(t)$ may be fit to a function of time and σ.

C.3 Modified Moment Closure

The moment closure technique of the previous section may be modified by changing variables to $k = f(\lambda)$ such that $p(\lambda, 0)$ is a normal distribution and expanding $f(\lambda)$ as a Taylor series about the mean $\bar{\lambda}$ [162, 160]. In this case, $f(\lambda) = e^{\lambda}$. This method works by moment approximations whose lowest order terms are a function of lower moments and are somewhat more robust to error than the moment closure technique of the previous section.

The system redefined in terms of mean and moments of λ are

$$\begin{aligned} g(t)/g(0) &= \int_0^\infty p(\lambda, t)dk \\ \bar{\lambda}(t) &= \int_0^\infty \lambda p(\lambda, t)dk \\ v(t) &= \int_0^\infty (\lambda - \bar{\lambda}(t))^2 p(\lambda, t)dk \\ m_3(t) &= \int_0^\infty (\lambda - \bar{\lambda}(t))^3 p(\lambda, t)dk \\ M_4(t) &= ... \end{aligned} \tag{C.9}$$

Since the first order decay rate is $k = f(\lambda)$, the distribution of original mass decays as

$$\frac{d}{dt}(p(\lambda, t)g(t)/g(0)) = f(\lambda)p(\lambda, t)g(t)/g(0). \tag{C.10}$$

Next, take the derivative of (C.9) using equation (C.10) and expand $f(\lambda)$ in equation (C.10) as a Taylor series approximation about the mean $\bar{\lambda}$.

After a bunch of work, the following set of moment closure differential equations appear [162],

$$\frac{d}{dt}(g(t)/g(0)) = g(t)/g(0)(f(\bar{\lambda}) + f_2(\bar{\lambda})v + f_3(\bar{\lambda})m_3 + ... + f_i(\bar{\lambda})m_i + ... \tag{C.11}$$

$$\frac{d}{dt}\bar{\lambda}(t) = f_1(\bar{\lambda})v + f_2(\bar{\lambda})m_3 + ... + f_{i-1}(\bar{\lambda})m_i + ... \tag{C.12}$$

$$\frac{d}{dt}m_j(t) = \sum_{i=1}^{\infty} f_i(\bar{\lambda})\{m_{i+j} + -m_i m_j - j m_{i+1} m_{j-1}\} \tag{C.13}$$

where $f_i(\lambda) = 1/i![\frac{d^i}{d\lambda^i} f(\lambda)]$ and $m_1 = 0$.

It is commonly thought that higher moments are negligible and the moment expansion can be closed with approximations based on lower moments [163,160]. This is somewhat counter-intuitive since moments of the normal distribution increase rapidly with moment number. These assumptions strongly depend on the function $f(\lambda)$ and is further discussed in section C.3.3.

C.3.1 A partial closure method

A typical moment closure technique is to track moments up to the equation for dv/dt and to "close" the equations with the 3rd and 4th moments of a normal distribution with variance $v(t)$ and neglect all higher moments. Therefore $m_3 = 0$ and $m_4 = 3v(t)^2$. Since m_4 is the lowest order non-zero term in dv/dt it must be included, however since it is a higher order term in dg/dt and $d\bar{\lambda}/dt$ one might consider neglecting it. This results in a 3 ODE approximation [160, 161]

$$\begin{aligned} \frac{d}{dt}g(t)/g(0) &\approx g(t)/g(0)\left(f(\bar{\lambda}) + \frac{1}{2}v(t)\frac{d^2f}{d\lambda}(\bar{\lambda})\right) \\ \frac{d}{dt}\bar{\lambda}(t) &\approx v(t)\frac{df}{d\lambda}(\bar{\lambda}) \\ \frac{d}{dt}v(t) &\approx v(t)^2\frac{d^2f}{d\lambda}(\bar{\lambda}) \end{aligned} \tag{C.14}$$

C.3.2 A full closure method

Since an approximation for M_4 is already being used in the equation dv/dt, why not close with M_4 in all equations especially if the term with M_4 may be large (as discussed in section C.3.3.)

$$\begin{aligned} \frac{d}{dt}g(t)/g(0) &\approx g(t)/g(0)\left(f(\bar{\lambda}) + \frac{1}{2}v(t)\frac{d^2f}{d\lambda}(\bar{\lambda}) + \frac{1}{8}v(t)^2\frac{d^4f}{d\lambda^4}(\bar{\lambda})\right) \\ \frac{d}{dt}\bar{\lambda}(t) &\approx v(t)\frac{df}{d\lambda}(\bar{\lambda}) + \frac{1}{2}v(t)^2\frac{df^3}{d\lambda}(\bar{\lambda}) \\ \frac{d}{dt}v(t) &\approx v(t)^2\frac{d^2f}{d\lambda}(\bar{\lambda}) \end{aligned} \tag{C.15}$$

Equations (C.14) and (C.15) are two of many ways the moments can be closed.

C.3.3 $f(\lambda) = -\exp(\lambda)$

Since $k = \exp(\lambda)$ and λ is initially normally distributed, k is lognormally distributed initially with mean of $\log k = \mu$ and standard deviation of $\log k = \sigma$. Rescaling time by e^{μ} eliminates the parameter μ from the system so the *initial* mean log rate $\mu = 0$ may be assumed in the remainder of this document.

Number of Terms Required

It is not obvious that higher moments are negligible since moments of the normal distribution increase rapidly with moment number and and also since $\frac{df}{d\lambda} = f(\lambda)$ remains constant for all terms. The size of each term in (C.11) is plotted in figure C-1 as a function of moment number for $\sigma = 3$. It seems that about 18 moments are required when $\sigma = 3$.

The number of moments required for a given sigma can be approximated by finding the moment number associated with the largest term as a function of sigma. By taking the derivative of the even moments m_i of a normal distribution divided by $i!$ and setting it equal to zero, the peak term is found. Figure C-2 shows that the peak moment number increases like σ^2. Since more terms are required after the peak, the number of terms required $\sim \sigma^2$. This makes solving problems with $\sigma > 1$ relatively cumbersome since the umber of differential equations in the system is the number of moments. Marine sediment datasets with $\sigma = 6$ would likely require accurate approximation of

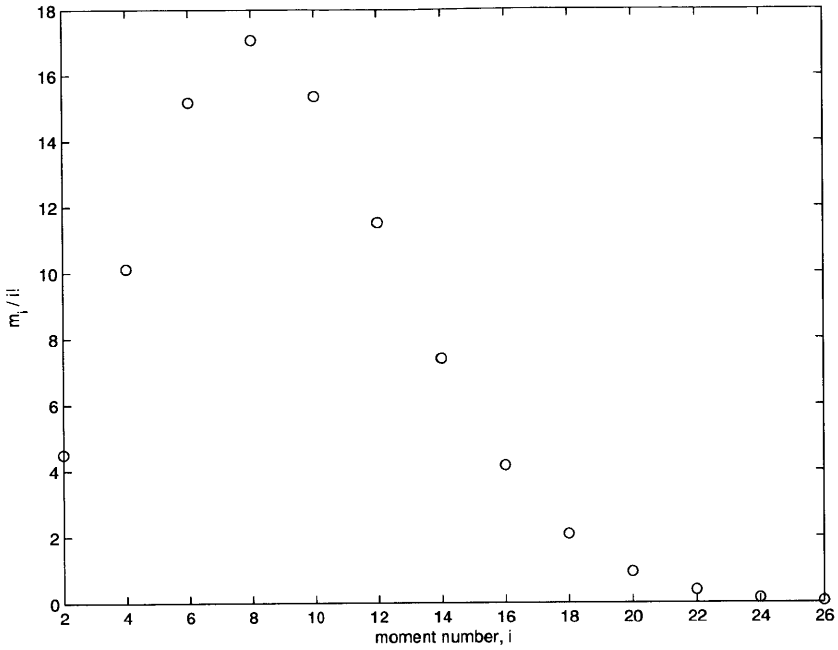

Figure C-1: Term sizes in first equation in (C.11) vs. moment number for a lognormal distribution with $\sigma = 3$ at $t = 0$. Terms higher than approximately the 18th moment may be neglected. $-f_i(\bar{\lambda}(0)) = e^{\mu} m_i/i!$

over 70 moments.

The feasibility of this moment closure technique is further investigated by solving (C.11) numerically. At each time step, the current mass remaining and initial variance is used to calculate the exact moments of a decaying lognormal distribution. These moments are fed back into the first equation of (C.11) to calculate dg/dt. The moment expansion decay can be compared to the true decay by finding the root mean square error (RMSE) of the approximation between $.95 > g/g(0) > .05$. Figure C-3 shows a plot of $\log_{10}$ RSME vs. number of terms in the expansion for four different values of σ. Approximations with $\log_{10}$ RMSE of around -2.5 or -3 are barely distinguishable from the true decay. When $\sigma = 3$, more than 14 terms are required for a good fit.

Furthermore, all approximations at sufficiently long time $g \ll .1$ diverge drastically from the true solution and easily become orders of magnitude in error. The approximation does not capture the precise dynamics of a decaying tail of a lognormal distribution.

3 ODE approximation with 4th moment in $v(t)$ equation

If $f(\lambda) = -\exp(\lambda)$ is used in equation (C.14) with initial conditions $\bar{\lambda}(0) = \mu$ and $v(0) = \sigma^2$, a rather nice closed form solution can be found for $\lambda(t)$, $v(t)$ and $g(t)$ [161]

$$g(t) = \frac{\sigma^{-2}\exp(-\sqrt{2e^{\mu}\sigma^2 t + 1})}{\exp(-\sigma^{-2})(2e^{\mu}\sigma^2 t + 1)^{1/4}}. \qquad \text{(C.16)}$$

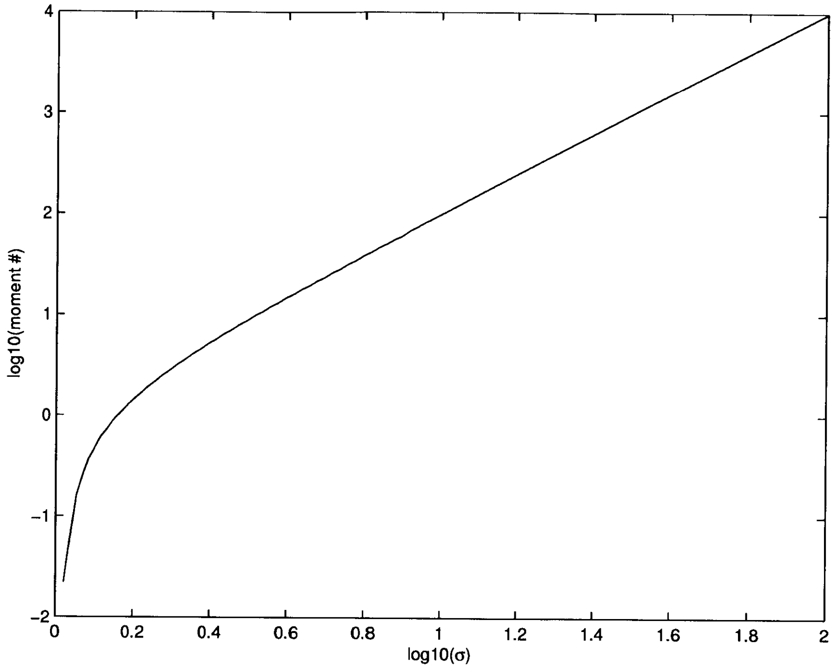

Figure C-2: Moment number of largest term in equation (C.11) vs $\log \sigma$. Largest term scales like σ^2, the number of terms required is likely twice the number of the largest term.

However as discussed in previous sections, this solution breaks down for $\sigma > 1$ as seen in figure C-4. Error becomes significantly noticeable at $\sigma = 1$ but by $\sigma = 3$ the approximation is about three natural orders of magnitude late predicting the start of decay and about three natural orders of magnitude early predicting when decay reaches 10% remaining. The error at early times is due to the poor approximation of the mean rate at $t = 0$ when $\sigma = 3$. The mean rate at $t = 0$ is $\bar{k}(0) = \exp(\sigma^2/2) = 90.0$ compared to the approximated mean rate $\bar{k}(0) = \sigma^2/2 = 4.5$ which differs by three natural orders of magnitude.

3 ODE approximation with 4th moment in all equations

Approximating $m_4(t) = 3v(t)^2$ in all three equations of (C.15) yields a significantly better fit to the true decay as shown in figure C-5. Unfortunately an analytical solution requires finding the root of an exponential integral according to `Maple`. This root can be found for asymptotically short and long times but even then, the asymptotically long time solution is still finding the root of $a(e^{\mu}t\sigma^2)Z^2 + e^Z$ which yields the ugly result $g(t) \sim f(\text{LambertW}(a(e^{\mu}t\sigma^2)))$.

Expanding each differential equation to include higher moments did not significantly improve the solution. In fact, on a log log plot, including higher moments resulted in much larger error at long times when $g \ll 0.1$.

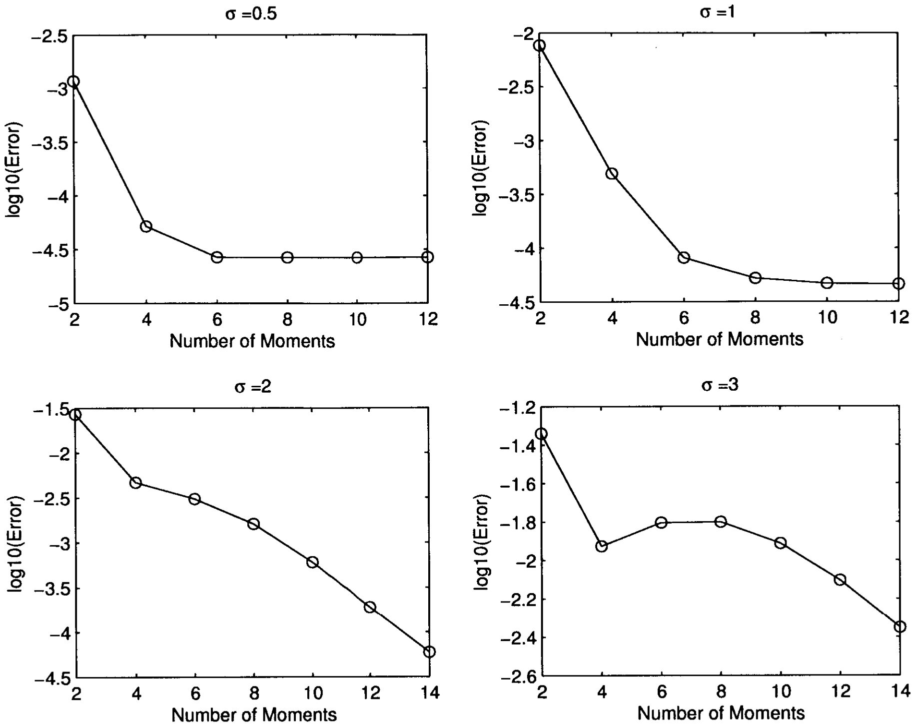

Figure C-3: log 10 RSME of numerical integration of Laplace transform of the lognormal vs. truncated series differential equation (C.11) truncated at the moment indicated on the X-axis. log 10 RSME errors of around -2.5, -3 are negligible. For $\sigma = 3$, about 16 to 18 terms are required agreeing with figures C-1 and C-2. Flattening of log 10 RSME around -4.5 is due to accuracy of numerical integration and/or numerical ode solver.

4 ODE approximation

Extending the moment closure technique to a 4 ODE system is not as straightforward. Applying equation (C.11) to the derivative of the 3rd moment results in $\frac{d}{dt}m_3(t) = 0$ when closing with the 4th moment $m_4(t) = 3v(t)^2$. Therefore a higher moment closure is required if any information is to be gained from the 4 ODE system.

This is where it gets tricky; closing with normally distributed higher moments assuming $m_5(t) = 0$ and $m_6 = 15v(t)^4$ in all 4 equations does provide a non-zero 4th equation but upon numerical investigation with an ode solver, it appears unstable.

Writing out the full equation for $\frac{d}{dt}m_3(t)$ using the same notation as in (C.11)

$$\frac{d}{dt}m_3(t) = f_1(\bar{\lambda})(m_4 - m_1 m_3 - 3v^2) + f_2(\bar{\lambda})(m_5 - v m_3 - 3v m_3) + f_3(\bar{\lambda})(m_6 - m_3^2 - 3v m_3) \quad \text{(C.17)}$$

then using closure $m_4 = 3v^2$ and eliminating $m_1 = 0$, all remaining terms are

$$\frac{d}{dt}m_3(t) = f_2(\bar{\lambda})(m_5 - 4v m_3) + f_3(\bar{\lambda})(m_6 - m_3^2 - 3v m_3). \quad \text{(C.18)}$$

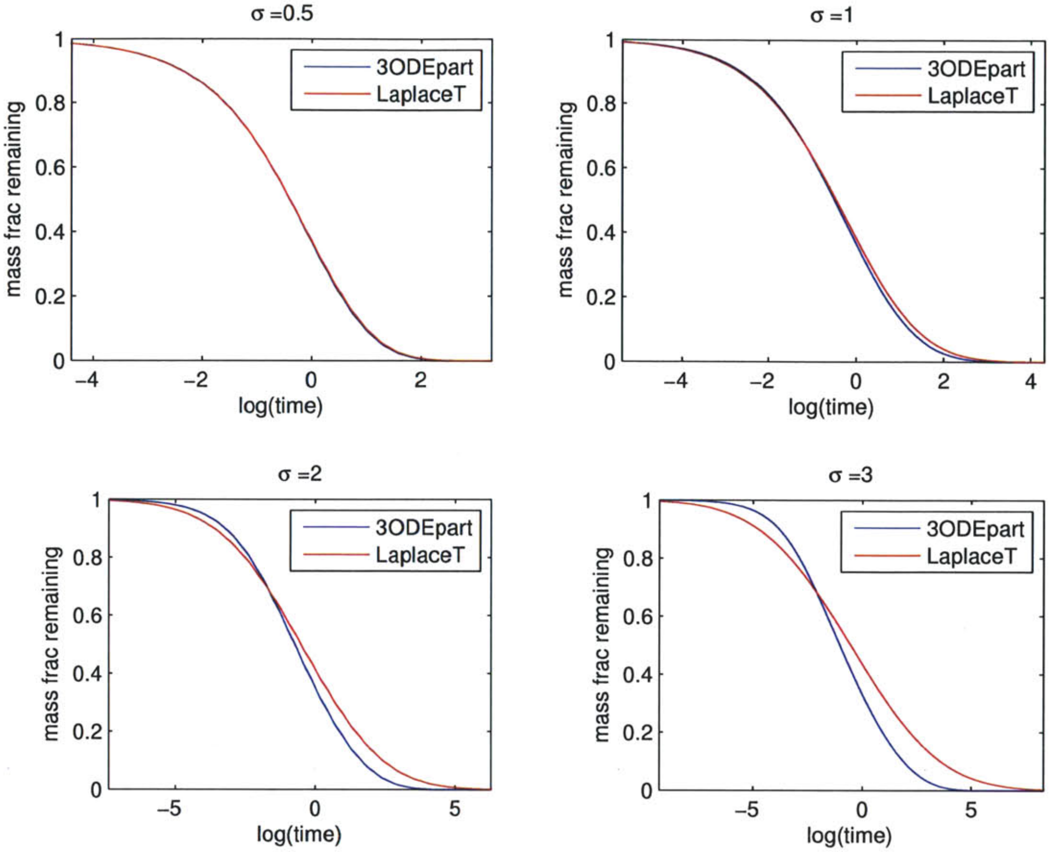

Figure C-4: Three ODEs with an approximation of the 4th moment only in the variance equation. This approximation is good for $\sigma < 1$.

The instability arises because we do not accurately keep track of the first term in (C.18) after $t = 0$. In (C.18), while $4vm_3$ grows, m_5 also grows and is likely of a similar magnitude. Therefore neglecting m_5 results in a positive pole and the instability develops. The same problem arises in the approximation for $\frac{d}{dt}v(t)$. Since the skew is being tracked, it should be possible to estimate m_5. Dimensions and intuition suggest that $m_5 = bm_3(t)v(t)$ where b is a constant factor. A good way to identify the appropriate b would be to numerically integrate the Laplace transform of the lognormal and plot m_5 vs. m_3v, however trial and error indicates that $m_5 = 12m_3v$ is an good choice. Closing all equations with estimates $m_4 = 3v^2, m_5 = 12m_3v, m_6 = 15v^4$ and substituting $\frac{d}{dt}f(\lambda) = -e^{\bar{\lambda}}$, a 4 ODE system to approximate the Laplace transform of the lognormal distribution is:

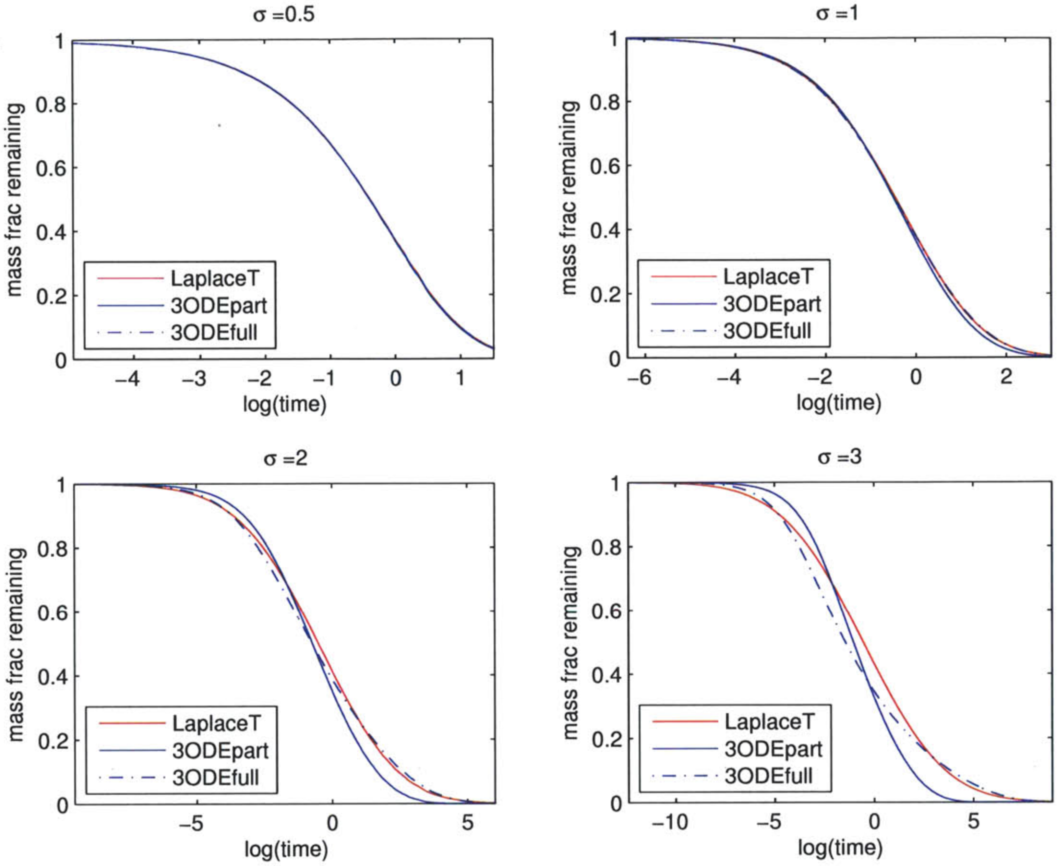

Figure C-5: Three ODEs with 4th moment approximation in all equations. This approximation is good for $\sigma < 1.5$. Some error is noticeable around $\sigma = 2$.

$$
\begin{aligned}
\frac{d}{dt}g(t)/g(0) &\approx -g(t)/g(0)e^{\bar{\lambda}}\left(1+\frac{1}{2}v+\frac{1}{6}m_3+\frac{1}{8}v^2+\frac{1}{120}m_5+\frac{1}{720}m_6\right) \\
\frac{d}{dt}\bar{\lambda}(t) &\approx -e^{\bar{\lambda}}\left(v+\frac{1}{2}m_3+\frac{1}{2}v^2+\frac{1}{24}m_5+\frac{1}{120}m_6\right) \\
\frac{d}{dt}v(t) &\approx -e^{\bar{\lambda}}\left(m_3+v^2+\frac{1}{6}(m_5-m_3v)+\frac{1}{24}(m_6-3v^3)\right) \\
\frac{d}{dt}m_3(t) &\approx -e^{\bar{\lambda}}\left(\frac{1}{2}(m_5-4vm_3)+\frac{1}{6}(m_6-m_3^2-9v^3)\right)
\end{aligned}
\tag{C.19}
$$

Making the substitutions $u = e^{\bar{\lambda}}$ as suggested by Bruggeman [161] and then $m_5 = 12m_3v$, $m_6 =$

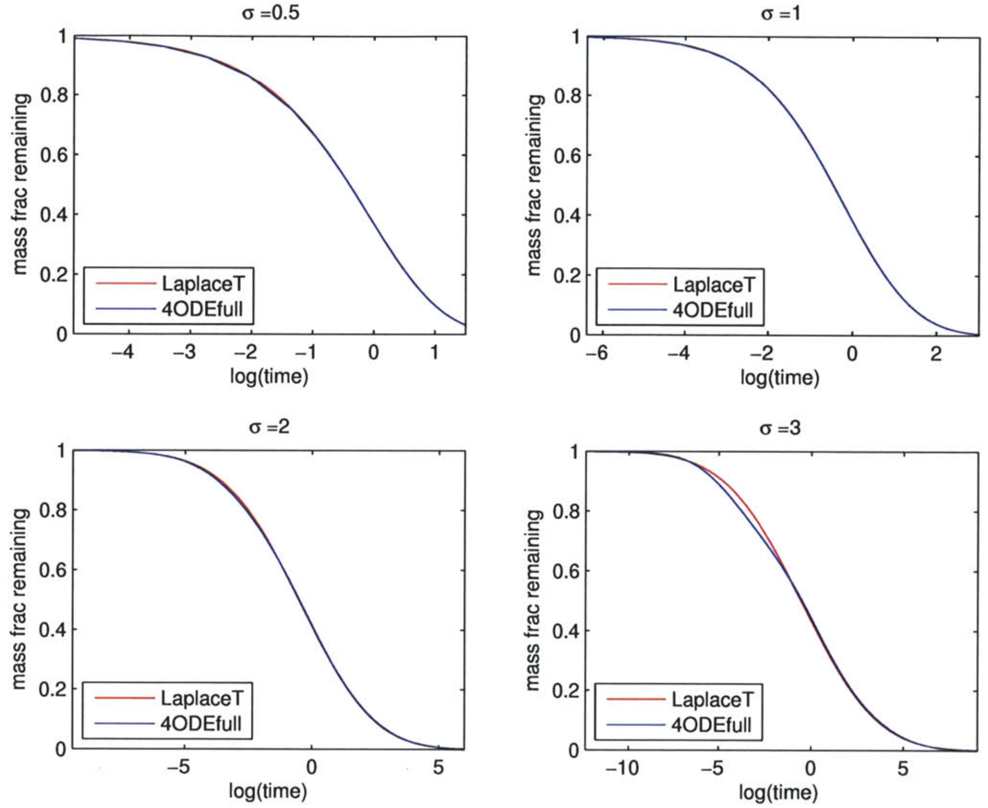

Figure C-6: 4 ODE approximation of the Laplace transform of the Lognormal. Tracks up to the 3rd moment about the mean. Accurate predictions between for $.99 > g > .01$ are obtained when $\sigma < 3$. 4th and 6th moments are closed using a normal distribution with variance $v(t)$ and the 5th moment is closed by $m_5(t) = 12v(t)m_3(t)$.

$15v^3$, the set of equations is

$$\begin{aligned}
\frac{d}{dt}g(t)/g(0) &\approx -g(t)/g(0)u\left(1+\frac{1}{2}v+\frac{1}{6}m_3+\frac{1}{8}v^2+\frac{1}{10}m_3v+\frac{15}{720}v^3\right) \\
\frac{d}{dt}u(t) &\approx -u^2\left(v+\frac{1}{2}m_3+\frac{1}{2}v^2+\frac{1}{2}m_3v+\frac{1}{8}v^3\right) \\
\frac{d}{dt}v(t) &\approx -u\left(m_3+v^2+\frac{11}{6}(m_3v)+\frac{1}{2}(v^3)\right) \\
\frac{d}{dt}m_3(t) &\approx -u\left(4m_3v+\frac{1}{6}(6v^3-m_3^2)\right)
\end{aligned} \tag{C.20}$$

The 4 ODE approximation provides a significant improvement over the 3 ODE approximation. As seen in Figure C-6, the 4 ODE approximation is good for $\sigma < 3$. Although the 6th moment closure decays too slowly for $g > .99$, the lag is hardly noticeable since the approximation quickly catches up to the true decay and even decays slightly too fast around $g \approx .8$. However, to get a tighter solution with RSME of -3, likely many more terms than the 6th moment are required.

Higher ODE approximation

Tracking the 4th moment was attempted but appropriate combinations of $v(t), m_3(t), m_4(t)$ in order to complete the 5th and 6th moments were not readily identified. Approximating 7th and/or 8th moments, and/or higher moments are required for a significant improvement over the 4 ODE system (C.20). The higher moment estimations can also be methodically done by monitoring the higher moments from a true lognormal decay and plotting them vs. lower moments to figure out the best combinations and factors of lower moments. However, as discussed in section C.3.2, accurate representation of around 18 moments is required for precise representation of the system when $\sigma = 3$.

C.4 Other Approximations to the Laplace Transform of the Lognormal

Since the characteristic function of the lognormal does not converge [164,150], there is not a readily available analytical approximation for moderate σ. A series solution for $g(t)$ exists but requires many terms even when $\sigma = 3$ and provides little physical insight to the decay [152]. Bruggeman's result (C.16) holds for approximately $\sigma < 1$ which is the most robust approximation I've seen. Previously, the best closed form analytical approximation was from Holgate [153]

$$g(t) \approx \frac{e^{-t}}{\sqrt{1+\sigma^2 t}} \exp\left(\frac{-\frac{1}{2}\sigma^2 t^2}{1+\sigma^2 t}\right) \tag{C.21}$$

which is decent for approximately $\sigma \leq .6$.

C.4.1 $\sigma \gg 1$ approximation

When $\sigma \gg 1$, the Laplace transform acts roughly like a step function filter since the actively decaying portion of the continuum is very narrow. After rescaling time by $\tau = e^{\mu}t$, $g(\tau)$ is simply a function of the cdf of the normal distribution

$$g(\tau) = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{-\log\tau + \gamma}{\sqrt{2}\sigma}\right)\right) \tag{C.22}$$

where log time is shifted by a small fraction $\approx \gamma$ (Euler's constant) due to the small absolute error in the filter assumption. Approximation (C.22) is good for $\sigma > 5$.

C.4.2 Conclusions

With Bruggeman's result (C.16) for $\sigma < 1$ and (C.22) for $\sigma > 5$ a compact analytical approximation for the Laplace transform of the lognormal when $1 < \sigma < 5$ remains to be found. Further application of the moment closure technique yields a set of 4 differential equations (C.20) which can be solved numerically and provide decent approximation for $\sigma < 3$.

Appendix D

Instantaneous Reaction Rate $K(t)$

This chapter investigates how the apparent first order reaction rate K changes as a function of time when rates are lognormally distributed.

The first order instantaneous rate is defined as $K = -\frac{dg}{dt}/g$ and is important because it is the apparent reaction rate at a time t. This rate characterizes how decay slows down with time. Instantaneous respiration rates are well known to decrease with time [36], [17], [103]. Middelburg described this relation by power law

$$K = .21t^{-.99}, \tag{D.1}$$

The relation (D.1) described a compilation of many experiments involving different types of organic matter in various locations over about ten orders of magnitude in time. The pre-factor in this relationship is still not understood.

Before discussing the scaling in (D.1), it is first helpful to consider the instantaneous reaction rate in a single reactive continuum data set. The definition of instantaneous rate reveals that

$$\begin{aligned} K(t) &= \int_0^\infty kp(k)e^{-kt}dk / \int_0^\infty p(k)e^{-kt}dk && \text{(D.2)} \\ &= \int_0^\infty kp(k;t)dk && \text{(D.3)} \\ &= \langle k(t) \rangle && \text{(D.4)} \end{aligned}$$

where $p(k;t) = p(k)e^{-kt} / \int_0^\infty p(k)e^{-kt}dk$ is the probability distribution of k remaining after time t and $\langle k(t) \rangle$ is the time dependent mean of the distribution $p(k;t)$. Therefore $K(t)$ is also the mean of $p(k;t)$.

D.1 $K(t)$ of Lognormally Distributed Rates

Figure shows a plot of $\log \mathcal{K}$ vs. $\log \tau$ where $\mathcal{K}(\tau) = K(\tau)/e^{\mu}$, $\tau = te^{\mu}$. This characteristic curve describes the instantaneous reaction rate for any time and μ while $\sigma = 1$. Characteristic rate

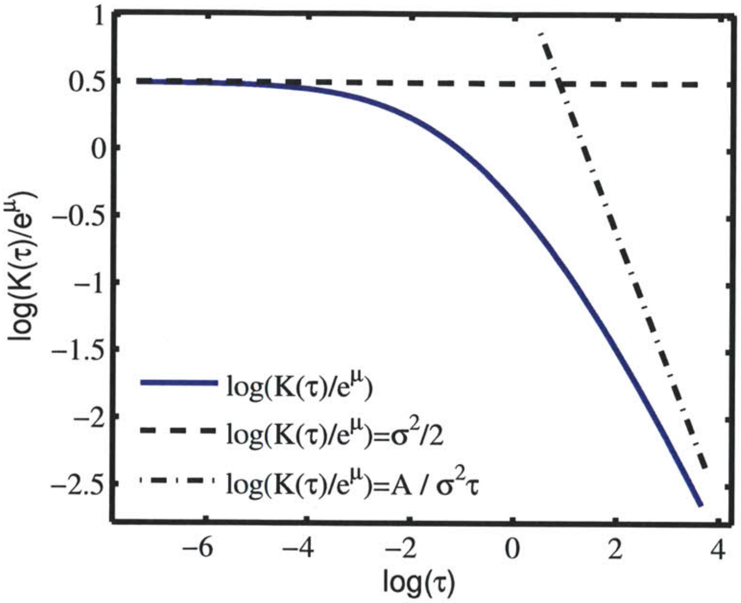

Figure D-1: $\log K(\tau)/e^{\mu}$ vs. $log(\tau)$ for a lognormal distribution with $\sigma = 1$. The horizontal asymptote as $t \to 0$ is simply the mean rate at $t = 0$ which is of course $K(0) = \mu + \sigma^2/2$. The main blue curve was calculated from the numerical integration of (D.2). The asymptote as $t \to \infty$ has parameter $A(10,1) = 3.85$. $A(\tau,\sigma)$ varies slowly with τ and σ. This asymptote is based on the mode of the distribution $p(k;t)$ which asymptotically approaches the mean of the distribution as $t \to \infty$ when $p(k;0)$ is lognormal.

curves for other values of σ are shown in figure D-2. The effect of sigma is easily seen. $\sigma = 0$ is a flat line of course as there is only one rate. Each have an asymptote of $\mathcal{K} = \sigma^2/2$ as $\tau \to 0$. The crossover point must also start earlier with increasing σ since the mean rate is obtained much more quickly. The $\tau \to \infty$ asymptote also decreases strongly with sigma since a larger fraction of slower rates reduce the instantaneous rate. To show this we first convert to a log distribution $\rho(\lambda;\tau)d\lambda = p(k;\tau)dk$ and $\lambda = log(k/e^{\mu})$ here. The mean of pdf $\rho(\lambda;\tau)$ will be approximated as the mode of $\rho(\lambda;\tau)$ since $\rho(\lambda;\tau)$ approaches a delta function as $\tau \to \infty$ and the mode approaches the mean.

$$\rho(\lambda;\tau) = \frac{e^{-\lambda^2/(2\sigma^2)}e^{-e^{\lambda}\tau}}{\int_{-\infty}^{\infty} e^{-\lambda^2/(2\sigma^2)}e^{-e^{\lambda}\tau}d\lambda} \tag{D.5}$$

Setting $\frac{dp}{d\lambda} = 0$, we find that the mode is located at

$$\lambda_{mode} = -\text{LambertW}(\tau\sigma^2) \tag{D.6}$$

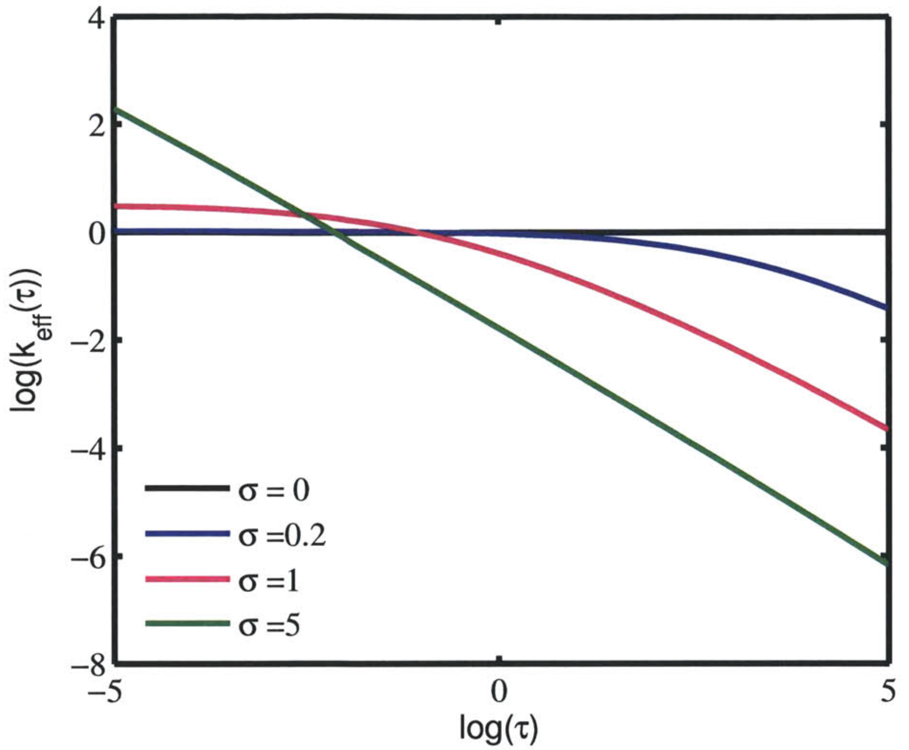

Figure D-2: $\log(K(t)/exp(\mu))$ vs. $log(\tau)$ for various σ. Four different values of sigma are shown.

The asymptotic expansion of the LambertW function can be used since we've already assumed that $\tau\sigma^2 \gg 1$.

$$\lambda_{mode} \approx -\log(\tau\sigma^2) + \log\log(\tau\sigma^2) - \log\log(t\sigma^2)/\log(t\sigma^2) \tag{D.7}$$

$$= \log(t\sigma^{2^{-1}}) + \log(\log(t\sigma^2)^{1-1/\log t\sigma^2}) \tag{D.8}$$

$$= -\log(t\sigma^2) + \log A(\tau,\sigma) \tag{D.9}$$

where A varies slowly with τ and σ. Because the mode of $p(\lambda;\tau)$ approaches the mean of $p(k;\tau)$ as $\tau \to \infty$

$$\langle k(\tau)\rangle/e^{\mu} \sim A(\tau,\sigma)/\sigma^2\tau \tag{D.10}$$

$$K(t) = \langle k(t)\rangle \sim A(\tau(t,\mu),\sigma)/\sigma^2 t \tag{D.11}$$

so the pre-factor in that asymptote varies directly with $1/\sigma^2$.

Thus for $\tau \gg 1/\langle k(0)\rangle$ distributions with larger σ will always have a slower $\mathcal{K}$. This happens physically because the effect of σ is to widen the distribution meaning that for a given change in time, a smaller fraction will be available to degrade as more of the distribution is further away from the active decay rates. Additionally for times $\tau \gg 1/\langle k(0)\rangle$, widening sigma will cause the distribution to be heavier on the slow side, which reduces the mean instantaneous rate $\mathcal{K}$.

Case of $\sigma \gg 1$

Due to the Laplace transform, the degree to which σ affects K depends on σ and must be calculated numerically for moderate σ. For $\sigma \gg 1$ however, there are closed form solutions of $g(t)$ and $\frac{dg(t)}{dt}$. Differentiating $g(t)$ (B.11) gives

$$\frac{dg(t)}{dt} = -\frac{e^{-1/2(-\ln(t)-\mu)^2/\sigma^2}}{\sqrt{2\pi}t\sigma}. \tag{D.12}$$

K scales with $1/\sigma$ as seen when dividing (D.12) by (B.11) to get

$$K(t) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma t}\frac{e^{-1/2(-\ln(t)-\mu)^2/\sigma^2}}{1+\mathrm{erf}((-\ln t-\mu)/(sqrt(2)\sigma))}. \tag{D.13}$$

The right hand product in (D.13) is ~ 1 for most of the decay resulting in

$$K(t) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma t} \tag{D.14}$$

when $\sigma \gg 1$ and $(-\ln(t)-\mu)^2 \ll \sigma^2$. Numerical investigation reveals that the approximation (D.14) is reasonable even if $(-\ln(t)-\mu)^2 \sim \sigma^2$.

D.2 Scaling $K(t)$ from Multiple Experiments

Alone, none of these figures are related to the Middleburg plot since the Middleburg result comes from an aggregate of studies. No single experiment can give such a large range of timescales. We therefore considering a few separate studies where the parameters are changing and compare $K(t)$. For simplicity, we first consider when decay dynamics are similar but occur on vastly different time scales hence μ varies but σ remains the same. As shown in figure D-3, decay dynamics are qualitatively similar but occur on vastly different timescales. Varying μ causes the shift $\log(\tau) = \log(t) + \mu$, $\log(\mathcal{K}) = \log(K) - \mu$ which results in the inverse time relation between K and t.

D.2.1 Effect of σ and sampling time on pre-factor A

The pre-factor A of the relation $K(t) = A/t$ varies with σ and the sampling interval. A varies with the sampling interval simply because earlier reactions occur at faster rates and later reactions at slower rates as seen in figure D-1. As mentioned in section D.1 σ effects the fraction of the distribution that is instantaneously decaying. Although the actively decaying matter at $\log \tau = 0$ may have $\log \mathcal{K} \sim 0$, a large fraction of the distribution may have $\mathcal{K} \ll 0$ and not be actively decaying, resulting in a mean rate $K < 0$.

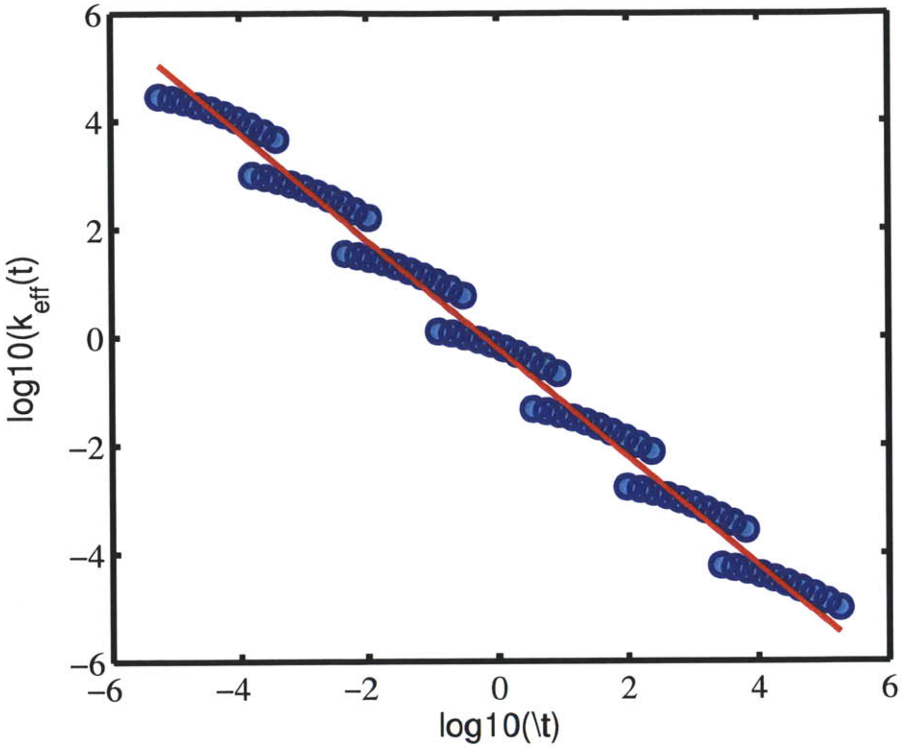

Figure D-3: $log(K)$ vs. $log(t)$ for various values of μ. Here, $\sigma = 1$ and the intercept A of the best fit curve with slope of -1 is $A = .61$.

Figure D-4 reveals the effect of σ on A. A diversity of rates in each sample with a standard deviation of around $\sigma \sim 3.5$ is consistent with Middelburg's $A \sim .2$. $2 < \sigma < 6$ results in $.1 < A < .4$. Middelburg's universal A indicates that there may be a characteristic diversity or rates or niches responsible for organic matter respiration. Figure D-4 was generated assuming that experiments were sampled symmetrically in $\log t$ about $\log t = \mu$. Of course $K(t)$ depends on sampling time, and any universal shift in the sampling time would change the pre-factor A. The combined effect of sampling time and sigma on pre-factor is seen in figure D-5. Measuring very early for each experiment will result in values of $A < 1$ since $t^{-1} > K(t)$. Measuring very late in each experiment will result in values of $A > 1$ since $t^{-1} < K(t)$ which corresponds to a situation when only a small mass remains which is still decaying at a mean rate which is faster than inverse time elapsed $1/t$. So $A \sim .2$ may arise if measurements are taken before the main timescales of decay or if there is a moderate variance in the decay rates. Both effects may contribute to the universal prefactor, however most experiments typically measure decay to more than 50% of the initial value indicating that a somewhat ubiquitous diversity of rates and niches may be controlling organic matter degradation dynamics.

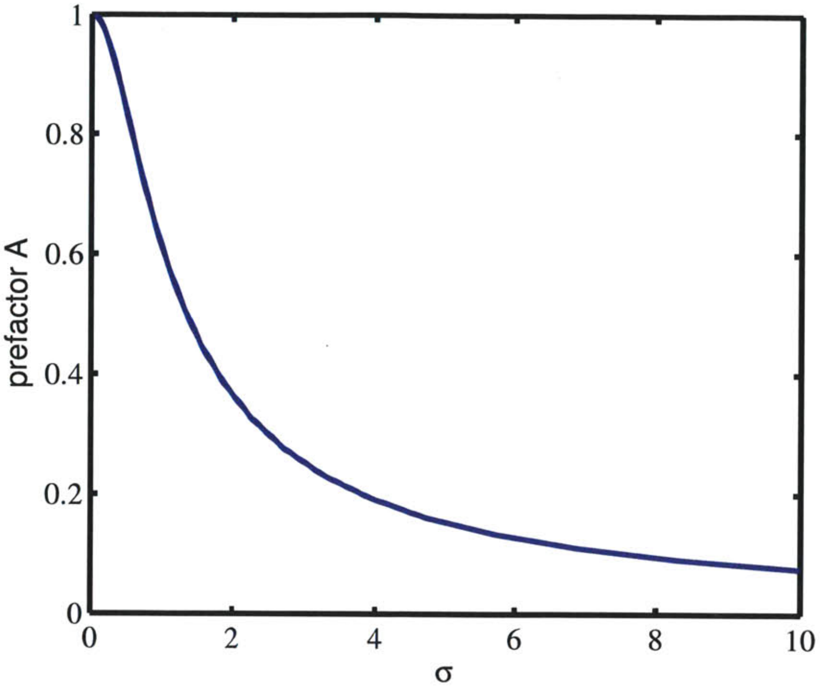

Figure D-4: For each value of σ, μ was varied over many orders of magnitude and the intercept A of a best fit curve of slope -1 was found. This figure compares the resulting intercept A vs. σ, and we see that moderate σ results in $A \sim .2$ However, this curve may be shifted if the decay is sampled earlier of later as seen in the next figure.

D.2.2 Characteristic values of σ predict $K(t)$

$K(t)$ is plotted in figure D-6 for many synthetic experiments having widely different values of μ and moderate σ. The data spans over 14 orders of magnitude. Each experiment in the figure had a $K(t)$ measured somewhere in the middle 86% of degradation and indicates that if datasets have distributions of σ, $2.5 < \sigma < 5$, then $K(t)$ is consistent with Middleburg's result $K(t) = .21t^{-.99}$. In other words, a moderate value of σ for decay on each timescale is sufficient to explain the occurrence of the pre-factor $A = .21$.

D.3 Mass Scaling of Heterogeneous Substrates

The calculation of K is dependent on the variance σ^2 of the rates of the sample which means that K in particular may depend on the sample size of the experiment if the variance increases with sample size. For example consider the following scenario.

A substrate of size $g_{1,0}$ and rate k_1 is added to another substrate of size $g_{2,0}$ with rate k_2. This

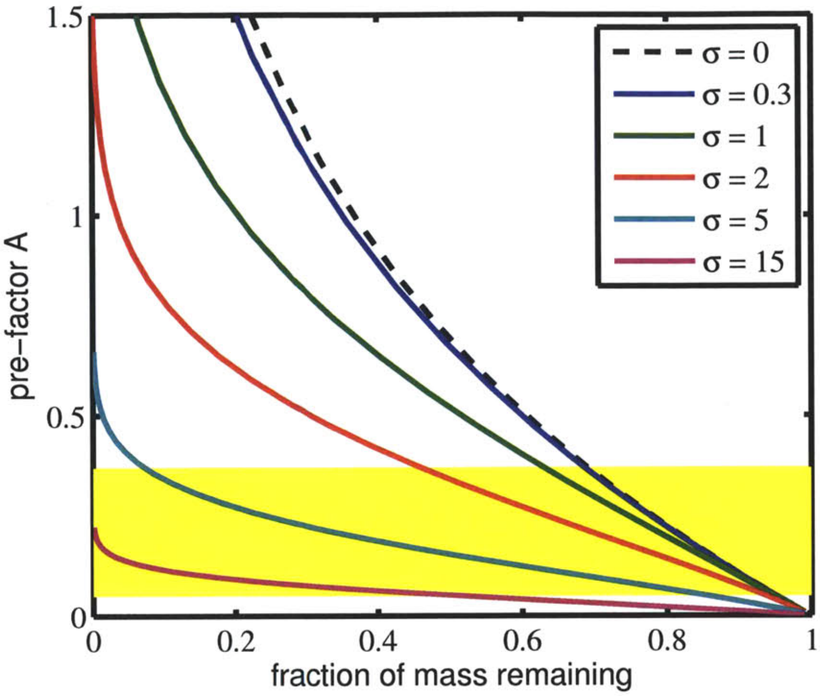

Figure D-5: This plot compares σ and mass remaining at the time of sampling to the intercept A. For a given mass fraction remaining and sigma, I show the corresponding the intercept A that would be found if μ were varied over a large timescale. For each value of μ there was only one sample taken at the desired mass fraction and σ in order to calculate A. It is easily seen here how different sampling locations may affect the intercept A. The yellow region is centered at Middelburg's intercept, $A = .21$ It is seen here that intercepts on the order of .2 are rather robust.

heterogeneous substrate is monitored and the following properties are calculated.

$$g(t) = g_{1,0}e^{-k_1 t} + g_{2,0}e^{-k_2 t} = g_1(t) + g_2(t) \tag{D.15}$$

$$\dot{g(t)} = -k_1 g_1(t) - k_2 g_2(t) \tag{D.16}$$

$$K(t) = \frac{k_1 g_1(t) + k_2 g_2(t)}{g(t)} \tag{D.17}$$

In the case of $k_1 = k_2$

$$\dot{g(t)} = -k_1 g(t) \tag{D.18}$$

$$K(t) = -k_1 \tag{D.19}$$

$K(t)$ is not mass invariant if the substrate is heterogeneous and additional mass can add new rates. The properties $g(t)$ and $\dot{g(t)}$ vary linearly when rates are added because superposition is used to calculate them.

One can imagine that if a smaller probe allowed investigation at the level of a tracheid (single

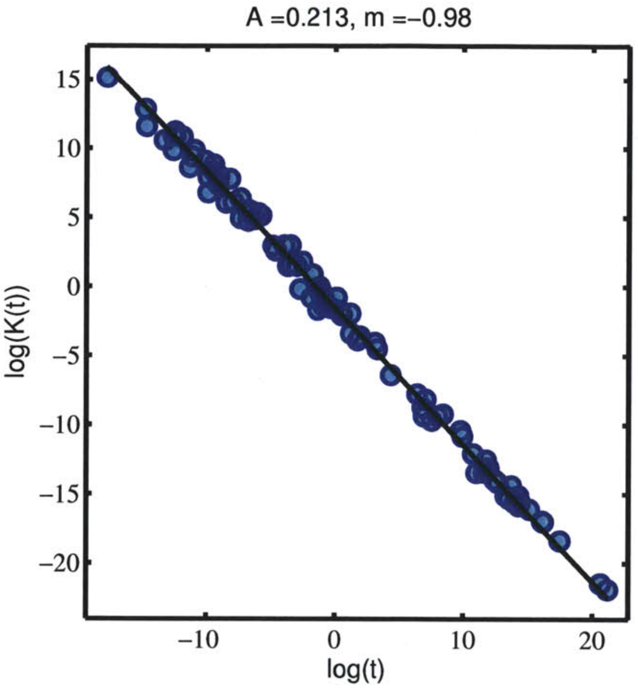

Figure D-6: $\ln K$ vs. $\ln t$. 100 synthetic data points are plotted, each calculated based on a random μ, σ, and mass fraction remaining g. g and σ were chosen over a uniform distribution with $.07 < g < .93$, $2.5 < \sigma < 5$ and μ was chosen over a normal distribution with mean of 0 and standard deviation of 15. The black line is Middleburg's instantaneous reaction rate, $K = .21t^{-.99}$. The RMA fit of this data yields $K(t) = .21t^{-.98}$.

cell in a fiber), to a lignin-hemicellulose matrix, to a hemicellulose polymer, different reaction rates would be observed as well as different residence times. An assembly of $K(t)$ curves based on measuring the decay of each polymer in a litter bag would exhibit the same qualitative trend as an assembly of tracheid curves but have a different prefactor A. Another example is increasing the variance of rates in a sample. For instance take the sum of all litter in the litter bags in the LIDET dataset or all of the sediment cores in the ocean drilling program. If we could probe total organic matter in surface sediments on an ocean scale, g is linear and would be a sum of individual measurements. However $\dot{g}/g$ from equation (D.2) is nonlinear and rescaling widely distributed fluxes $\dot{g}(t)$ by a huge g therefore changes K from the bulk by orders of magnitude compared to plotting the sequence of $K(t)$ in the individual cores. This is precisely why σ has an inverse relationship with K.

For example, consider a collection of a large number of samples each having lognormally distributed rates with parameters μ and σ as in figure D-6. Each sample has a random μ which is widely distributed and all samples have identical σ. Calculating K for each sample at random sampling times and plotting $\log K$ vs $\log t$ results roughly in the power law relation as in figure D-6. Alternatively, consider the case where the same samples are collected but each sample cannot be distinguished and are monitored as one large mass. According to conditional probability, this mass has a distribution of rates

$$p(\log k) = \int_{-\infty}^{\infty} \rho(\log k|\mu) b(\mu) d\mu \tag{D.20}$$

where μ is distributed randomly with pdf $b(\mu)$ and $\rho(\log k|\mu)$ is the aforementioned lognormal distribution of rates given the parameter μ. Assuming μ is normally distributed, $p(\log k)$ is difficult to calculate analytically. However some physical insight regarding diffusion sheds light on the subject. The solution to the diffusion equation with initial condition $b(\mu)$ is the convolution

$$p(\log k) = \int_{-\infty}^{\infty} (\sqrt{2\pi}\sqrt{2t})^{-1} e^{\frac{\log k - \mu^2}{4t}} b(\mu) d\mu \tag{D.21}$$

which is not necessarily easily integrated. However, if $b(\mu)$ is normally distributed with mean m and variance s^2, then the solution to (D.21) is easily found since $b(\mu)$ itself is a solution to the diffusion equation at an intermediate time t_1 with initial condition of a delta function at $\log k = m$. Upon diffusion of $b(\mu)$ for an additional time t_2, the normal distribution $p(\log k$ is found. Using the relation between time and variance S^2, we find that $p(\log(k))$ has variance

$$S^2 = \sigma^2 + s^2 \tag{D.22}$$

and the reactivity distribution is

$$p(\log k) = \mathcal{N}(m, S) \tag{D.23}$$

Sample distinguishability therefore is very important when calculating K but not important when calculating g or $\dot{g}$. When samples can be distinguished, the decays are power laws with coefficient $A(\sigma)$ determined from figure D-4. When a continuum of samples are indistinguishable and the entire mass is monitored, it degrades like one large reactive continuum. For the case of μ distributed lognormally, $p(\log k)$ is normal and K behaves as described in this chapter. If $S = \sqrt{\sigma^2 + s^2} \gg 1$, eq (B.11) can be used and K is easily approximated analytically as $K(t) \approx \sqrt{2}(\sqrt{\pi}\sigma t)^{-1}$ for most of decay. Although distinguishable and indistinguishable samples both lead to similar inverse power dependence K on t, the pre-factor $\sqrt{2/\pi}\sigma^{-1}$ is significantly different than A.

In summary, when calculating any quantity from degradation studies it is important to realize that the decays measured reflect the distributions of rates in that sample and a larger sample may have a wider distribution of rates. While the effect of sample size on $g(t)$ is a straightforward superposition, sample size and wider rates may strongly affect certain calculated quantities such as $K(t)$. Therefore the meaning of $K(t)$ should be carefully interpreted.

Appendix E

Pathway Estimation

This chapter can be considered the discrete distribution analog to chapter 4. Similarly chapter 3 is the discrete distribution analog to chapter 2. Here, I seek the discrete exit rate function which best describes the decay data constrained such that $g \geq 0$ and $dg/dt \leq 0$. In other words, I find a multi-pool model where pools are allowed to be negative. Specifically, I seek the multi-pool solution with the fewest number of pools that describes the decay data but not the noise. Then I provide an approach for estimating the most likely underlying network matrix **A** based on the discrete exit rate function. In this manner, the best fitting first order decay pathway can be estimated.

By breaking the problem of pathway identification into two parts: 1) finding the best fitting discrete exit rate function and 2) estimating plausible network architectures from that exit rate function, the inverse problem for pathway estimation now has a reduced set of free parameters. Calculating the discrete exit rate function identifies the total number of pools in the network, provides constraints on the connectivity of the network matrix, and also provides a good initial guess at the transfer rates associated with those pools. A brute force inversion of the data for all possible network configurations **A** would be much more difficult and computationally intensive.

Before continuing, I note that that this approach succumbs to the same pitfalls seen for the multi-pool models of chapter 3. The discrete exit rate functions are highly sensitive to noise. In fact, in this case they are so sensitive to noise that it is even hard to identify the best fitting discrete exit rate function, because adding more pools allows better fits to the noise even though solutions are constrained. I therefore find this method, like the multi-pool method inappropriate for modeling the LIDET datasets which contain a large amount of noise. However, I include this section because it may have applications for other systems believed to be well described by continuous time Markov processes with a finite number of states, such as in pharmacokinetics or systems biology.

E.1 Identifying the Discrete Exit Rate Function

As derived in chapter 4, a decomposition network with N pools

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \mathbf{A}\mathbf{x}$$

has the impulse response

$$g_N = \sum_{i}^{N} r_i e^{k_i t} \tag{E.1}$$

I now seek to identify solution with the smallest number of pools N that well describes the data. To do so, I vary N from 1 to D^1. Then, for each value of N, I calculate the best fitting parameters r_i and k_i by solving the optimization problem

$$\min(\|\hat{g} - g_N\|) \tag{E.2}$$

subject to the constraint

$$dg(t)/dt \leq 0 \tag{E.3}$$

for all time t. This optimization problem is carried out in Matlab using the Matlab optimization algorithm `fmincon.m`. The global minimum is commonly found when randomly choosing 20 initial conditions in the solution space and solving (E.2) subject to (E.3) for each one. Implementing the constraint (E.3) for all time t however is somewhat tricky.

E.1.1 Constraint $dg/dt \leq 0$

To implement this constraint, first note that

$$\frac{\mathrm{d}g_N}{\mathrm{d}t} = \sum_{i}^{N} -k_i r_i e^{k_i t} \tag{E.4}$$

I then make the change of variables $w = e^{-t}$ in equation (E.1).

$$\frac{\mathrm{d}g_N}{\mathrm{d}t} = \sum_{i}^{N} -r_i k_i w^{k_i}. \tag{E.5}$$

Because the positive time axis collapses to the interval $[0, 1]$ on the w axis, the polynomial (E.5) is constrained to have no zero crossings over the interval $0 \leq w \leq 1$. This constraint is somewhat efficiently implemented by using the function `fminbnd.m` in Matlab to search for the minimum of

---

[1] The maximum number of pools D is arbitrary but can be chosen in various ways: 1) Choose the D such that a D pool model has the same number of parameters as data points,2) Choose a value of D such that any higher number of pools results in no further significant reduction in residual error.

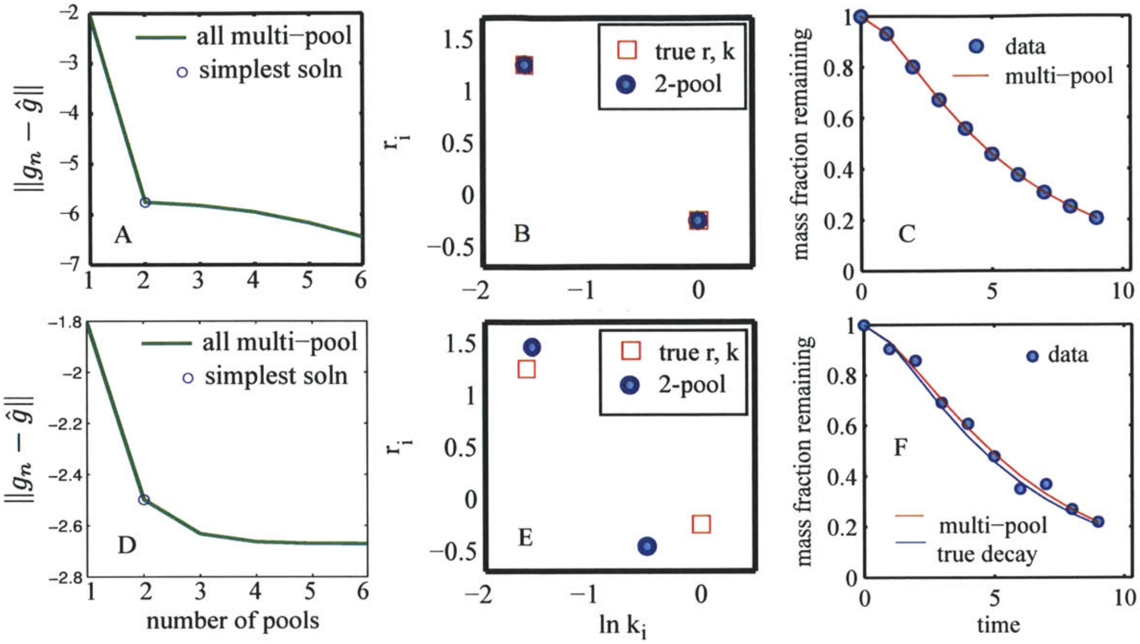

Figure E-1: Multi-pool inversion, 2 pool example. A: L-curve for optimal solutions from 1-6 pool models when data has low noise. B: parameters associated with the 2 pool solution when data has low noise. C: decay associated with 2 pool example with low noise and best fitting two pool solution. D: L-curve for optimal solutions from 1-6 pool models when data has low noise. E: parameters associated with the 2 pool solutions when data has moderate noise. F: decay associated with 2 pool example with moderate noise and best fitting 2 pool solution.

$-dg_N/dt$ over the interval $[0,1]$. The constraint used in the optimization algorithm is then

$$\min_w \sum_i^N r_i k_i w^{k_i} <= 0, \quad 0 \leq w \leq 1. \tag{E.6}$$

E.1.2 Example problems

Two-State System

Before looking at real data, consider the two state serial degradation pathway of Figure 4-1 with parameters $k_1 = 1$, $k_2 = .2$ and $f = 1$. The associated exit rates of this system are the eigenvalues $\lambda_1 = k_1$ and $\lambda_2 = k_2$. The weights r_i are found from equation (4.44), $r_1 = -.25$ and $r_2 = 1.25$. The decay g is then calculated and Gaussian noise with zero mean is added to the decay.

In the first example I consider only a tiny amount of Gaussian noise having a standard deviation $\sigma = 0.001$. I then vary the number of pools N from 1 to 6 and solve (E.2) subject to the constraint (E.6) in order to find the optimal parameters r_i and k_i for each pool model. The results are shown in Figures E-1ABC.

Figure E-1A shows a plot of the residual error vs number of pools for the six models considered.

The residual error decreases with pool number, but the improvement in residual error becomes marginal with higher pool number. Similar to the L-curve analysis of chapter 3, the corner of this curve represents an optimal trade off between solution complexity and fitting the data. The corner correctly identifies the correct 2 pool solution which is shown in Figure E-1B. The data and predicted decay from the model is shown in Figure E-1C.

I then repeat this experiment with larger noise, $\sigma = 0.03$, which is a characteristic for some LIDET datasets. The results of the experiment with larger noise are shown in Figures E-1DEF. The corner of the curve shown in Figure D with now less defined, although it still suggests a two pool model. The parameters of the best fitting two pool model is shown in E and no longer match the parameters of the underlying system, but remains somewhat close. The difference in parameters occurs because these parameters fit the noise in the data better than the parameters of the underlying system. Other instances of the data at this noise level result in even a less well defined corner, and the parameters r_i and k_i deviate significantly from the true system. The decay of the true system, the data, and the prediction are shown in Figure E-1F.

Three-State System

I now consider an example pathway containing three pools in series as shown in Figure E-2. The

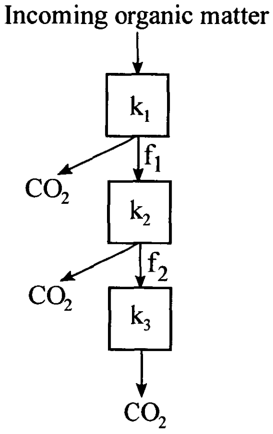

Figure E-2: Three serial pool example.

parameters $k_1 = 1.3$, $k_2 = 0.4$, $k_3 = 0.05$, $f_1 = 1.0$, $f_2 = 0.5$ result in the eigenvalues $\lambda_1 = k_1$, $\lambda_2 = k_2$, $\lambda_3 = k_3$ and weights $r_1 = -0.2133; r_2 = 0.6190; r_3 = 0.5943$.

Figure E-3 shows the results of the inversion approach when both a small $\sigma = 0.001$ and a moderate $\sigma = 0.03$ amount of noise is added to the decay data from this system. Figure E-3A shows a well defined corner identifying a three pool solution when the noise is small. However, even for this small amount of noise, the values of r_i and k_i do not exactly match the parameters of the underlying system as shown in Figure E-3B. Figure E-3C however shows that these parameters also fit the data unmistakably well. Figure E-3D shows that the L-curve becomes much more ambiguous when the noise is moderate and the choice of the solution is not obvious. Figure E shows the parameters associated with both the 3 and 4 pool solutions compared to the true parameters of the system. The parameters are off from the true solution because they describe the noisy data better than the underlying system does.

However, the r_i and k_i of the best fitting solutions also provide clues about the possible pathways from which they derive. The three pool solution has two pools with similar rates k_i but large positive and negative values of r_i. This is consistent with the behavior of pools in series with no loss as described in section 4.4.2. The third pool has a different rate and a positive weight $r_i < 1$. Based on this behavior and my intuition after working with these systems, I suggest this pathway contains effectively two pools in series and one in parallel, as shown in Figure E-3G. There may be possibility for transfer from the serial pools to the separate third pool. This hypothesis is tested in the next section on a real decay system with similar behavior. Similarly, the 4 pool model contains three pools with large fluctuating positive and negative r_i at similar rates k_i. For the same reasons, I suggest that these parameters are consistent with a possible pathway consisting of by three pools in series and one in parallel shown in Figure E-3H.

E.1.3 LIDET decay data

Figure E-4 shows the results of the multi pool inversion of the LIDET dataset depicted in Figure E-4C. Given the large amount of noise, the corner of the L-curve is ambiguous and it is not clear what the simplest solution is that captures the trend of the data. An alternative approach for choosing the model with the fewest number of pools that that best fits the data is to employ the Aikake Information Criterion (AIC) [131]. AIC is a procedure for comparing models and choosing the one with the fewest parameters that best fits the data. It does this by employing a metric that compares the number of degrees of freedom in the model to the maximum likelihood of obtaining the observed error between prediction and model. I suggest this approach as a possible alternative if this inverse method is to be applied in the future.

Since it is not clear which solution is the simplest that best describes the data, the parameters from the optimal 3 and 5 pool are shown in figure E-4B. However, it is rather clear from the decay predictions in Figure E-4C that the 5 pool model contains added dynamics which aren't necessarily needed to describe the data, such as the inflection point before the first data point and the slowdown at $t = 5$ and then speed up in decay around $t = 10$. Although I don't prove it here, I believe decay with such inflection points can only derive from networks with feedback. The three pool model has the same properties as seen in Figure E-3E: two pools with similar rates but large positive and negative weights with the third pool at a different rate and a positive weight fraction less than one. Therefore, for the same reasons I suggest that the discrete exit rate function possibly derives from the pathway shown in Figure E-4D.

To verify that the architecture of the pathway shown in Figure E-4D is plausible, I set up the

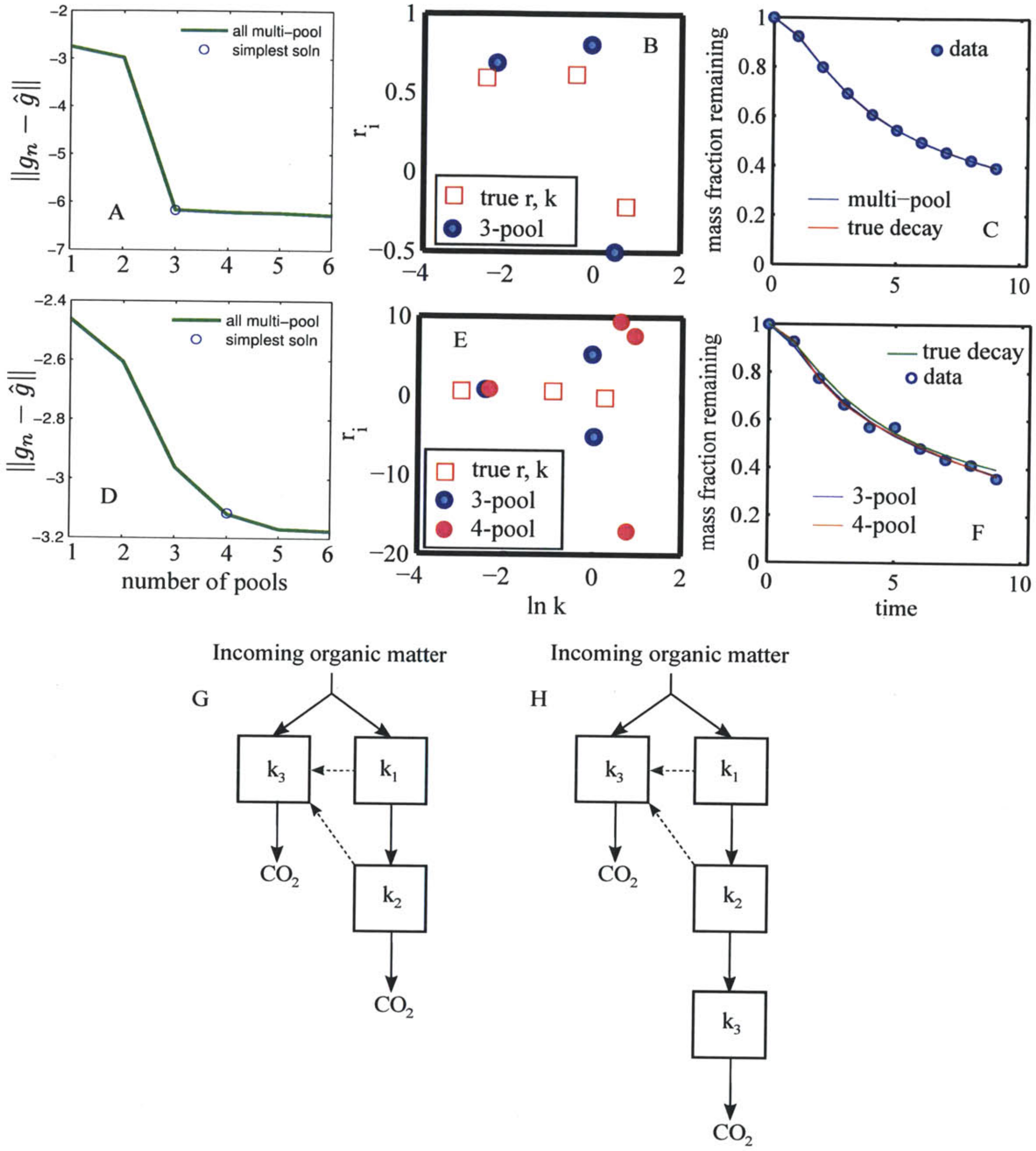

Figure E-3: Multi-pool inversion, 3 pool example. A: L-curve for optimal solutions from 1-6 pool models when data has low noise. B: parameters associated with the 3 pool solution when data has low noise. C: decay associated with 3 pool example with low noise and best fitting three pool solution. D: L-curve for optimal solutions from 1-6 pool models when data has moderate noise. E: parameters associated with the 3 and 4 pool solutions when data has moderate noise. F: decay associated with 3 pool example with moderate noise and best fitting 3 and 4 pool solutions. G: suggested pathway for 3 pool solution. H: Suggested pathway for 4 pool solution.

linear system matrix **A**, having parameters a, f_1, and f_2. Because the system has no feedback, **A** is triangular and the eigenvalues (the k_1) are the diagonals of **A**. I then solve an optimization

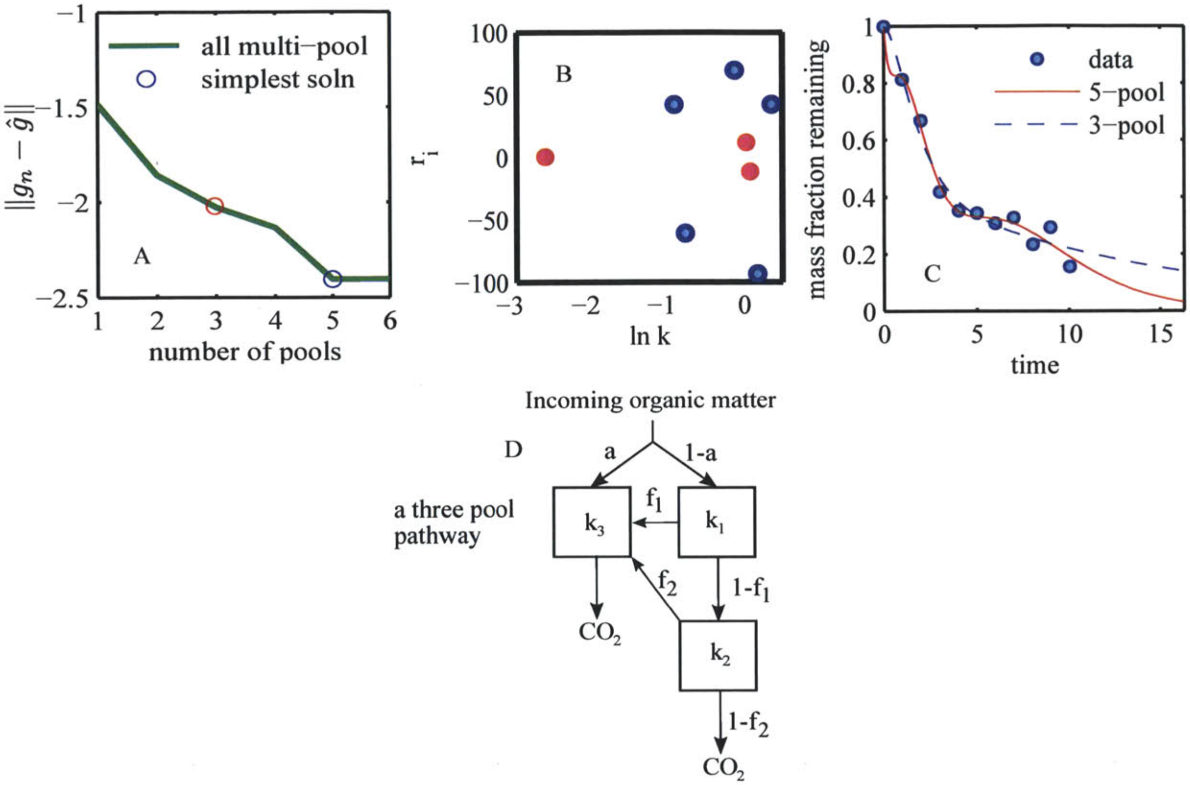

Figure E-4: Multi-Pool inverse of a LIDET dataset. A: L-curve for optimal solutions from 1-6 pool models. B: parameters associated with the 3 pool and 5 pool solutions. C: decay associated with the LIDET dataset and best fitting three pool solution and five pool solutions. D: Possible pathway architecture of 3 pool solution.

problem

$$\min_{a,f_1,f_2} \|\mathbf{r} - \mathbf{S}\mathbf{U}^{-1}\mathbf{p}_o\|, \tag{E.7}$$

where $\mathbf{r}$ is the vector of observed values r_i, $\mathbf{U}$ contains the eigenvectors of $\mathbf{A}$ (see eq. (4.48)), $\mathbf{S}$ is the diagonal matrix whose elements are the sum of the eigenvectors (see eq. (4.20)), and $\mathbf{p}_0 = [1-a,\ 0,\ a]^{\mathrm{T}}$ is the initial state partition vector (see eq. (4.13)). The optimization problem determines the values of a, f_1, and f_2 which correspond to the observed weights r_i. I then check the values to see if they violate the constraints $0 < a < 1$, $0 < f_1 < 1$, $0 < f_2 < 1$. I solve (E.7) using the algorithm `lsqnonlin.m` in Matlab. The values of the parameters for this particular pathway are $a = 0.026$, $f_1 = 0.22$ and $f_2 = 0.22$, all of which satisfy the constraints. Of course the pathway shown in Figure E-4D is just one of many possible pathways, as the three states can be connected in many more configurations, each of which may have more than 3 parameters, resulting in a large space of possible pathways.

All of the pools in the five pool network on the other hand have weights with large positive and negative values and eigenvalues k_i spread out, rather than clumped together. It is hard to guess at a no-feedback network configuration which will result in the observed weights and rates. Of the few lower triangular network configurations that I tried, none of them resulted in parameters f_{ij} and p_0 which satisfied the constraints of a Markov process or compartmental model. These parameters likely result from a feedback network with additional free parameters K_i along the diagonal of $\mathbf{A}$, as $k_i \neq K_i$.

An approach for more generally identifying all possible pathways is given in the next section.

E.2 Pathway Estimation from the Multi-pool Solution

Because the inverse method does not contain full information about the eigenvectors, the pathway matrix $\mathbf{A}$ cannot be directly recreated from it. There are an infinite number of systems $\mathbf{A}$ which have the same decay parameters r_i and k_i. For each system architecture, or each configuration of a connectivity matrix, there may be many more free parameters than the estimated multi pool parameters. This results in a volume of solutions in parameter space. However, this volume shrinks somewhat when we consider the constraints on decay systems $\mathbf{A}$. Namely these are constraints on connections from pool i to pool j, f_{ij}:

$$0 \leq f_{ij} \leq 1 \tag{E.8}$$

$$\sum_i f_{ij} \leq 1 \tag{E.9}$$

and constraints on the initial partition vector $\mathbf{p}_0$ with elements p_i:

$$0 \leq p_i \leq 1 \tag{E.10}$$

$$\sum_i p_i = 1. \tag{E.11}$$

Given these constraints, I suggest two brute force approaches for identifying likely pathways associated with the data. However, more sophisticated methods of system identification exist in control theory [165, 166].

E.2.1 Evaluating $\mathbf{p}_0^T \mathbf{r}^{-1} = \mathbf{U}\mathbf{S}^{-1}$

This approach proceeds in the following manner:

1. Choose plausible network architectures based on any additional constraints or physical processes governing the system. If all architectures are plausible, consider all possible configurations of the connectivity matrix. A possible connectivity matrix for an 3 pool network

is

$$\begin{bmatrix} x & 0 & 1 \\ 1 & x & 1 \\ 0 & 1 & x \end{bmatrix} \tag{E.12}$$

It is clearly seen that there are 2^{N^2-N} different possible connectivity matrices for N pools. When there are no connections (all except the diagonals are zero), this is simply the case of a parallel system with a partition vector whose indices are the weights r_i.

2. For each connectivity matrix, give the connections the values $k_i f_{ij}$ and calculate the eigenvector matrix U and the eigenvector sum matrix S.

3. Let the partition vector $\mathbf{p}_0$ have arbitrary elements p_{0i}. Calculate the inverse $\mathbf{p}_0\mathbf{r}^{-1}$.

4. This part requires a lot of algebra, but it should be possible to inspect the each element of $\mathbf{p}_0\mathbf{r}^{-1} = \mathbf{U}\mathbf{S}^{-1}$ to determine whether the constraints (E.8) and (E.10) are violated. If one of the elements violates the constraint, this configuration has no solution.

5. Estimate the intersection of the solution spaces for each element. This is the solution space for that configuration. If the likelihood of each parameter p_i and f_{ij} are known, it should be possible to estimate the likelihood of this pathway architecture.

6. Repeat previous steps for next pathway architecture.

7. With this information it should be possible to estimate the most likely pathway architectures.

E.2.2 Broad search optimization

This approach is perhaps even more brutal and obvious. For each combination of connectivity matrix of the system $\mathbf{A}$ and partition vector $\mathbf{p}_0$ calculate the optimal parameters of $\mathbf{A}$ by minimizing $\mathbf{r} - \mathbf{S}\mathbf{U}^{-1}\mathbf{p}_0$ based subject to the constraints (E.8) and (E.10). There are likely an infinite number of solutions for many of the configurations. Re-solving the optimization problem for various initial conditions may help to identify whether there is or is not a feasible solution. The number of different system configurations includes the different combinations of connectivity matrices and different permutations of non zero values in the partition vector $\mathbf{p}_0$, since this vector effectively sets the incoming 'edges' to the network.

This approach rapidly becomes cumbersome as the number of optimization problems scales with N as $2^{N^2-N}2^N$ where the first term is for different connectivity of the A matrix and the second is for different connectives of the $\mathbf{p}_0$ vector.

E.3 Discussion

The inversion method of section E.1 well identifies the underlying exit rates and their weights from a decay time series with little noise. However, the presence of noise seems to destroy much information about the details of the underlying network. This is why I advocate the regularized inversion of the exit rate function discussed in section 4. The AIC could possibly be used to identify the simplest multi-pool solution in the presence of noise if the system is truly believed to be governed by a network of a few first order transformations. However, the r_i and k_i for the chosen model will also be strongly dependent on the noise in the data. Further work is required however before being able to robustly identify architecture of the most probable pathway. Other system estimation methods [165] from control theory should be considered.

Appendix F

Noise in Decay Data

I showed in chapter 3 that the variance in the estimated rate distributions $\rho(\ln k)$ may be due to noise in the data, inherent heterogeneity in decay rates, or both. In this chapter I investigate noise from both situations in order better understand their effect on the variance of $\rho(\ln k)$.

F.1 Signal to Noise Ratio of the Decay Datasets

Before contemplating the origins of the noise in litterbag studies, I first discuss the magnitude of noise inherent to the data from the LIDET study.

The signal to noise ratio (SNR) of the LIDET data can be identified by taking the the standard deviation of the raw data at each removal time. Typically four litter bags were removed at a time for each substrate/site combination. The average percent mass remaining of those few bags is used as the data g. The signal is the mean of all the estimates of that data point g and the standard deviation of the noise of the signal is calculated according to the central limit theorem.

$$\sigma_s = \sigma_d/\sqrt{N} \tag{F.1}$$

where N is the number of litter bags averaged at the time of removal, and σ_d is standard deviation of the four raw data points at each removal time. The definition SNR $= g/\sigma_s$ is used therefore giving estimate of the SNR for each data point of the experiments we considered in the study. A histogram of the SNR of 4-bag data points is shown in figure F-1. Thus on average the SNR of LIDET is about 10 and may be much larger or smaller for some datasets. However, this amount of noise is large enough to warrant regularization.

Next I disuss the origins of the noise associated with F-1 so we can better interpret the estimated rate distributions $\rho(\ln k)$.

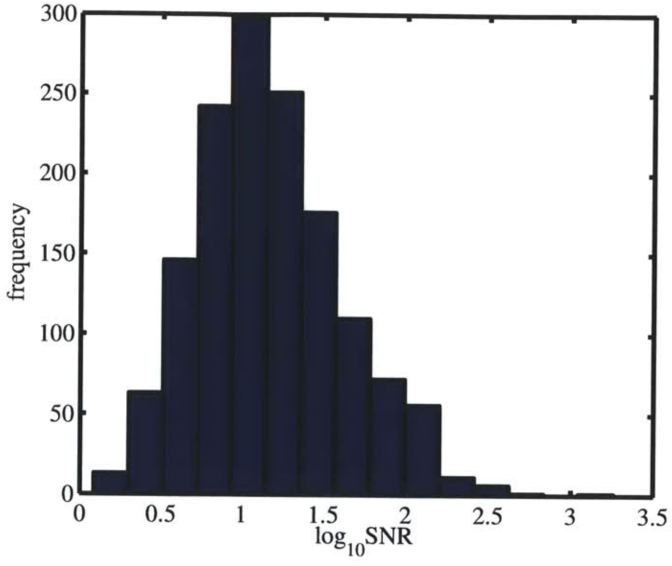

Figure F-1: histogram of SNR from the LIDET dataset. Only data points g_i averaged from four non-flagged bags are shown. 1454 out of 2022 data points were averaged from four bags, 1766 were averaged from at least three bags. The mean of $\log_{10} SNR$ in this histogram is 1.16.

F.2 Ill-posed Problem

Field-measured litter decay $g_{exp}(t)$ consists of a signal plus noise.

$$g_{exp}(t) = g(t) + e(t) \tag{F.2}$$

where g_{exp} is the mass measured in the field, g is the true mass, and e is the noise.

Standard modeling methods seek to minimize the error between prediction and experiment

$$\|g_{exp} - g_p\| \tag{F.3}$$

where g_p is the predicted mass. Assuming a linear model $g_p = \mathbf{A}p$ where p are the parameters of the model, the solution to the least squares problem (equation (F.3)) is

$$(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{g_{exp}} = \mathbf{p} \tag{F.4}$$

Plugging in eq. (F.2) into (F.4)

$$(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{g} + (\mathbf{A^T}\mathbf{A})^{-\mathbf{1}}\mathbf{A^T}\mathbf{e} = \mathbf{p} \tag{F.5}$$

A goal of regularization is to fit the signal and not the noise in problems which are ill-posed, meaning $\mathbf{A}^T\mathbf{A}$ is an ill-conditioned matrix containing singular values $\ll 1$.

However, because of the way litter decay is monitored, each datapoint represents a unique sample in a unique location. The high variability of nutrients in the same litter type likely results in different rates present from litterbag to litterbag. Therefore it is hard to know exactly what is the noise, since the variation in the the "true" decay rates of the data points likely vary with the "true" decay rates associated with each litter bag. This means the true signal is a combination of all of the decay rates.

Even though there is noise amongst the litterbags, some datasets still appear to be described best with only one rate. Noise from these bags may have nothing to do with variations of the actual decays between bags.

F.3 Noise from Random Error

Consider the situation where decay is exponential with rate k_0 Replace A in equation (F.5) with the time vector t, p with the rate parameter $-k$, and g with $\log g$.

$$||\mathbf{t}||^{-2}\mathbf{t}^T \log \mathbf{g} + ||\mathbf{t}||^{-2}\mathbf{t}^T\mathbf{e} = -k \tag{F.6}$$

$$||\mathbf{t}||^{-2}\mathbf{t}^T(-k_0\mathbf{t}) + ||\mathbf{t}||^{-2}\mathbf{t}^T\mathbf{e} = -k \tag{F.7}$$

$$-k_0 + \frac{\mathbf{e}^T\mathbf{t}}{||\mathbf{t}||^2} = -k \tag{F.8}$$

Here we see that k is a random variable which is the sum of the random error weighted by time. Thus if error is absolute with mean zero and variance σ_e^2, k should be distributed about k_0 with a variance of $\sum \sigma_e^2 \frac{t_i}{||t||^2}$.

If error is relative and denoted by random variable r_i then

$$e_i = \log \frac{g_i + r_i g_i}{g_i} \tag{F.9}$$

$$e_i = \log 1 + r_i \tag{F.10}$$

substituting equation (F.10) into equation (F.8) reveals that k is the sum of the log of random variables. The central limit theorem suggests that k should be lognormal, and we see from Figure (1) that even a sum of ten datapoint yields a distribution of $\log k$ which looks Gaussian.

Regularization was applied to a number of the same trials of $g_{exp}(t)$ to identify the distribution of reaction rates $\rho(\log \kappa)$ associated with $g_{exp}(t)$. These trials indicated that $\rho(\log \kappa)$ is Gaussian, however in all trials the standard deviation of $\rho(\log \kappa)$ was about ten times larger than the distribution of $\log k$ in Figure (1).

F.4 Noise Due to Differences In Litterbags

We now consider the case where the error comes from slight differences in the characteristics of litterbags, likely variations in nutrient composition from one bag to another. This may result in

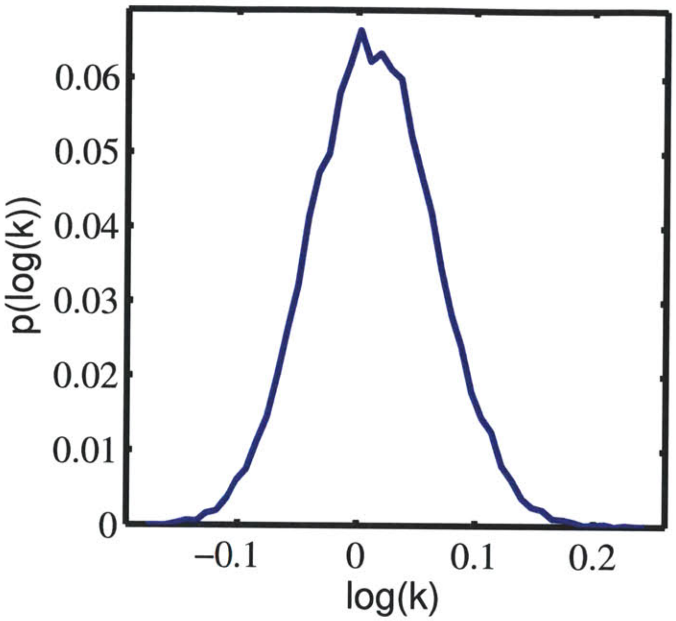

Figure F-2: Distribution of $\log k$ from 20000 trials. For each trial experimental data was created by setting $g_{exp}(t) = g(t) + rg(t)$ at ten equally spaced time points from 0 to $2k_0$, were relative error r is a normally distributed random variable with a mean zero and standard deviation = .2. k was calculated for each trial using equations (F.10) and (6).

different degradation speeds of different bags. For now, assume each has its own rate κ which is sampled from a lognormal distribution $\rho(\log \kappa)$ having median rate k_0 and standard deviation s.

The experimental decay $g_{exp}(t)$ in this case is

$$\log g_{exp_i} = -\kappa_i t_i \tag{F.11}$$

An example of this decay is shown in Figure F-3. The data g_{exp} from Figure F-3 is inverted using Tikhonov regularization and the resulting $\rho(\log \kappa)$ is shown in Figure F-3. Although there are only ten datapoints and the noise is heavy, the regularization *surprisingly identifies* the distribution of κ responsible for the experimental data. This is remarkable especially since decay g_{exp} appears proceed in an almost linear fashion in Figure F-3. It means that if the noise in the decay dataset is due to fluctuations of the distribution from bag to bag, the regularization technique can roughly identify the inherent distribution associated with all litter bags sprinkled about that forest. In this respect, the regularization technique may be able to distinguish noise resulting from heterogeneity from other background noise in the dataset.

Furthermore, the mean g_{exp} from Monte Carlo simulations of many trials collapses to the true decay $g(t)$ as the number of trials increases, since each datapoint is representative of the entire distribution and the error of the mean goes to zero. Of course each litterbag likely contains a distribution of rates and is most likely not the worst case scenario of equation (F.11) where a bag

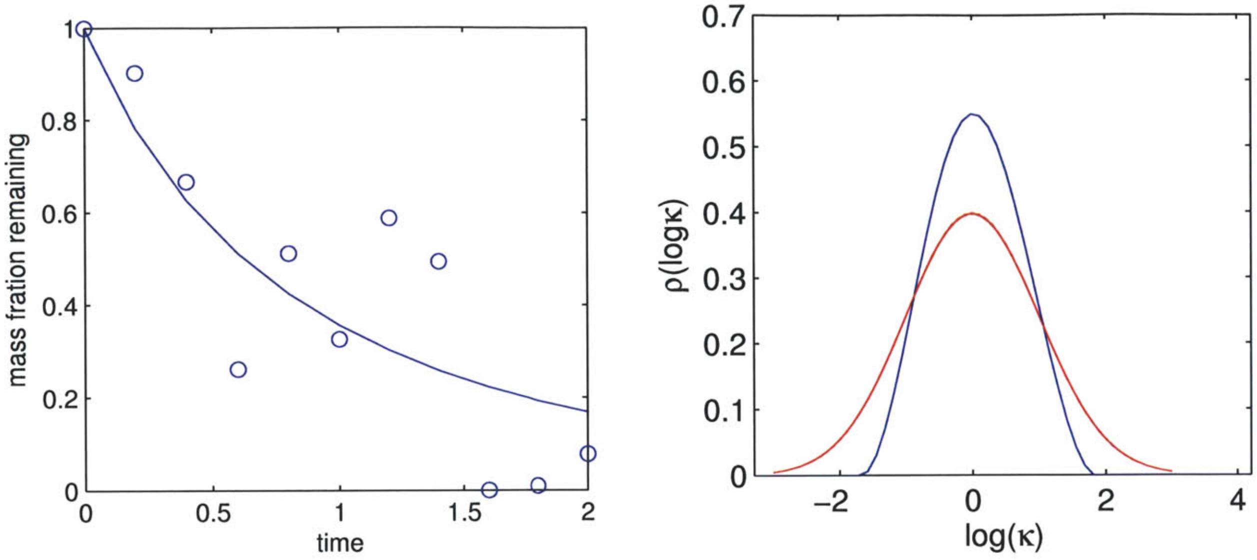

Figure F-3: Left: g_{exp} vs t. Here $\log g_{exp_i} = -\kappa_i t_i$ where κ is a random variable pulled from $\rho(\log\kappa)$ having median rate $k_0 = 0$ and standard deviation $s = 1$. Line is forward transform of inversion result $\rho(\log\kappa)$ shown in the figure right. Right: blue line is $\rho(\log\kappa)$, the inversion of $g_{exp}(t)$ of figure left. $\rho(\log\kappa)$ is shockingly similar to the lognormal distribution (red line) from which the random κ is pulled for each of the datapoints. The data actually reminds me more of the error in the LIDET datasets than other synthetic data I've played with; it looks like real LIDET decay data.

consists of particles with identical rates, however, as bags become more mixed and representative of the underlying distribution, g_{exp} appears to have less noise and approach $g(t)$. Of course the decay rates are likely continuous in a macroscopic litter bag, but the same principles apply; although the continuous rate distribution may change from bag to bag resulting in what appears to be noise, the regularization technique attempts to find the underlying distribution associated with that sub sample from which the bags are taken.

This also leads to the related case of $\sigma = 0$. There may be only a small variance of rates in a dataset which doesn't necessarily show up when other "noise" is significant and perhaps time dependent. This may be why we cannot tell the difference between the fit of solutions having $\sigma = 0$ and $\sigma \ll 1$ as the decay $g(t)$ look exponential for both. Decay with $\sigma = 0$ may have coarse heterogeneity which may vary from bag to bag, while datasets showing a slowdown may have heterogeneity within each bag. The regularization method appears to characterize both scales.

Unfortunately, because the data in Figure F-3 looks linear, fitting the lognormal to the data g_{exp} (as described in section 3.5) yields the parameters $\mu = .009$ and $\sigma = 0$. Fitting just the parameters μ and σ cuts right through the sample noise, whereas the regularization characterizes the noise as well as resulting from a variation in rates.

F.5 Fitting a Single Exponential to Distributed-Rate Decay

Here I consider using least squares to fit a single exponential decay rate k to the decay g_{exp} resulting from a continuous superposition of rates. The distribution of fitted values of k as a function of random bag rates κ is

$$k = \frac{\kappa^T \mathbf{t}}{||\mathbf{t}||^2} \tag{F.12}$$

Equation (F.12) results from canceling the terms $||\mathbf{t}||^{-2}\mathbf{t}^T \log \mathbf{g}$ in equation (F.6). Here, k a weighted sum of random variables κ.

It can be noted here that equation (F.6) provides the least squares fit of an exponential to the distributed rate decay with zero noise: $k^* = ||\mathbf{t}||^{-2}\mathbf{t}^T \log \mathbf{g}$.

F.6 Noise in g from Underlying Rate Distribution

Here I continue with the idea that the scatter in the decay curves results from sampling the underlying the distribution of rates.

F.6.1 Individual litter bags with homogeneous rates

Like in section F.4, I assume that litter bags sample only one rate, but that rate varies from bag to bag. Next, I suggest an approach to estimate the underlying rate distribution given this assumption.

Check distribution of error posteriori

One idea is to solve the inversion problem as usual with Tikhonov regularization making all of the standard assumptions, then tabulate the k_i or $\ln g_i$ from each litter bag (40 measurements, not averaged) and compare it to the distribution of estimated k_i or $\ln g_i/t$ from the regularized distribution.

This method will likely indicate a significant difference in the two distributions if $\sigma > 1$ since $\log(g(t))$ becomes nonlinear and the k_i from larger t will be biased towards slower values as more bags have $g \approx 0$ and no signal at large t.

Identify distribution of rates apriori

Instead of assuming a model and back checking its error to see if it works, we can instead identify the underlying rate distribution from the measured $k_i = \ln g_i/t_i$ by finding an optimal distribution that maximizes the likelihood of sampling all of the k_i.

Of course this method will also fail for the same reason as in the previous section when $\log(g(t))$ is nonlinear as k_i then appears to be a function of time, so there is a bias in the distribution at later times. In other words, the g_i are not independent: g_i is related more to its neighbors than points further away.

Regardless, continuing with this idea, I use a different regularization scheme which is based on maximizing the likelihood of obtaining the data $g_i(t)$ or $k_i(t)$, rather than minimizing $\|g_{exp}(t) - g(t)\|$. Continuing with the assumption of one rate per bag, the probability of sampling all k_i from the underlying distribution is

$$P = \prod_{i=1}^{n} p(k_i)\Delta k \quad \log P = \sum \log p(k_i) + \log \Delta k \tag{F.13}$$

This assumes there is slight error in the measurement of k_i or $\log k_i$ so the probability of sampling withing the measurement plus error is the slice of the pdf $p(k_i)\Delta k$. Assuming the Δk is the same for each datapoint means I can simply ignore the proportionality constant and maximize

$$\log P = \sum \log p(k_i). \tag{F.14}$$

Equivalently, I can find the maximum likelihood of the distribution in log-space $\rho(\ln k)$. $\rho(\ln k)$ can be discretized on the $\ln k$ axis and the maximization problem (F.14) can be re-expressed as

$$\min\{A \log \rho\}. \tag{F.15}$$

A is a diagonal matrix with elements a_i equal to the weight of each $\ln k_i$. The two nearest neighbors to the measured k_i are linearly interpolated and are assigned weights which sum to one. You can imagine the case where each $\log k_i$ landed on a grid point, with one repeating. Each a_i associated with a measured $\log k_i$ would be set to $a_i = 1$ while the repeated measurement would be set to $a_i = 2$ since we want to maximize the probability of getting that $\log k_i$ twice or $p(\log k_i)^2$. a_i is zeros for remaining values of $\log k$ on the grid which were not sampled.

The solution of course is the distribution $\log p = \frac{a}{n\Delta}$ where n is the number of litter bags and Δ is the spacing of discretization of $\log k$. However this solution is rough. To find a more likely solution of underlying rates, we use regularization. I tried many forms of the regularization operator including two forms of entropy: $\max \sum \log \rho$ and $\max \sum \log(-\rho \log \rho)$, minimizing the norm $\log \rho^T \log \rho$, and minimizing the first and second derivative.

Some of these regularization operators look linear in *logp* however the constraint is that $\sum e^{\log \rho} = 1$, which is highly nonlinear, requiring use of the `fmincon.m` function.

With the initial guess to be the distribution $\mathbf{a}/n$ The entropy and min norm operators all seemed to take the initial guess at the distribution and just flatten it to a loguniform. The derivative operator seemed to want to make a tent, and the second derivative made a Gaussian like distribution in $p(k)$ for the corner of the L-curve. Likelihood values $\log P$ varied from -6 to -60, with an optimally smooth solution having a value of -10. I compare this result with the solution from our standard regularization method from the previous chapters, where the residual error is minimized.

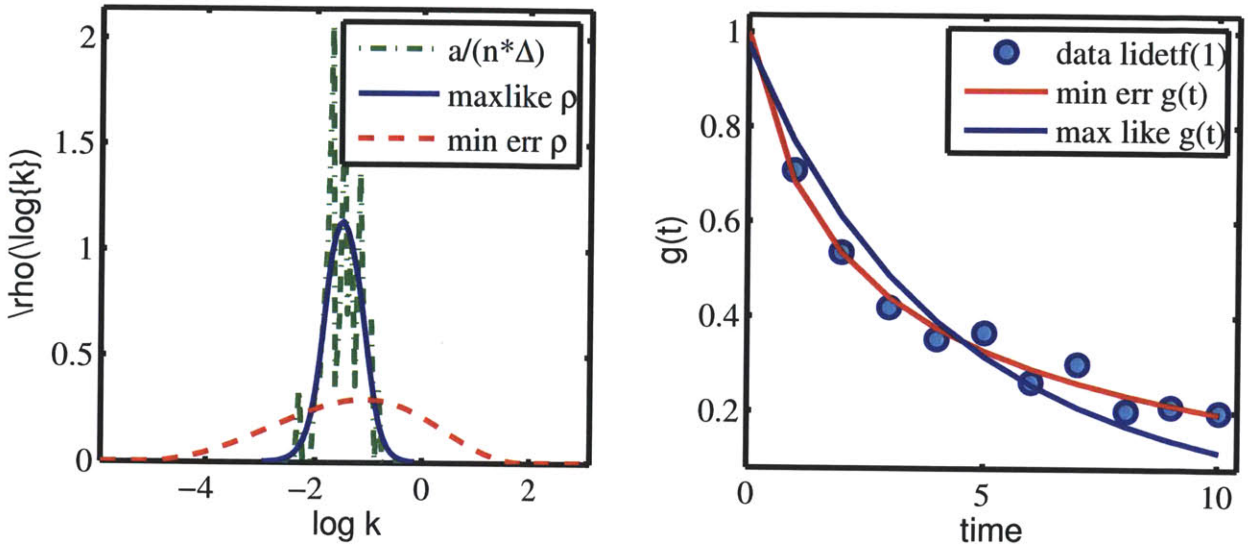

Figure F-4: Left: inversion $\rho(k)$ from maximum likelihood estimate with a second derivative regularization operator (blue). Also shown is distribution of k_i in green and the Tikhonov regularization error in blue. Right: data from LIDET dataset with fit from Tikhonov regularization (red) and maximum likelihood of obtaining k_i (blue).

The maximum likelihood distribution is narrower than the one obtained by the regularization technique with about the same center. Plotting the prediction of $g(t)$ from the maximum likelihood solution, revealed poor fit, which underestimated the data at long times (due to not estimating the bias in k_i at long times). This clearly means each of the bags for this dataset have heterogeneous rates.

F.6.2 Each litter bag contains many rates

At this point there are two paths to pursue if we still believe error is from the underlying distribution. We could assume each litter bag contains the master distribution of rates, just shifted and stretched. Or we could assume each bag contains many random samples from the underlying distribution.

Proceeding with the latter, we can write down the following equations.

$$g_p = 1/N \sum_i^N e^{-k_i t} \tag{F.16}$$

N is the number of components in each bag sampled from the distribution and is $\gg 1$. Note that each element $X_i = e^{-k_i t}$ in g_p is distributed as

$$\begin{aligned} p(X) &= p(k)\left|\frac{dk}{de^{-kt}}\right| \\ &= p(k(X))\frac{1}{e^{-kt}t} \end{aligned} \tag{F.17}$$

And g_p is the sum of random variables with the distribution (F.17).

From here we will try to maximize the likelihood of obtaining g_{exp} from a system like equation (F.13). Because $N \gg 1$, the random variable g_p should be normally distributed.

g_p has the mean

$$\langle e^{-kt}\rangle = \int_0^\infty p(k)e^{-kt}dk \tag{F.18}$$

$$= A\mathbf{p}. \tag{F.19}$$

while the variance of $g_p(t)$ about its mean is

$$\langle (e^{kt} - \langle e^{-kt}\rangle)^2\rangle = \int_0^\infty \left(e^{-kt} - \langle e^{-kt}\rangle\right)^2 p(k)dk \tag{F.20}$$

$$= A_2\mathbf{p} \tag{F.21}$$

The pdf of obtaining $g_{exp}(t)$ at a given time is

$$P(g(t)) = \mathcal{N}(A\mathbf{p}, A_2\mathbf{p}). \tag{F.22}$$

where $\mathcal{N}$ denotes the normal distribution. The problem is then to find $\mathbf{p}$ by maximizing the likelihood of getting the experimental measurements $g_{exp}(t)$.

$$\max \sum_i^n \log P(g_{exp,i}) \tag{F.23}$$

This represents a separate approach for estimating decomposition heterogeneity inherent in all litter bags. Its an alternative to the regularized inverse Laplace transform of the data.

Appendix G

Microbial Wood shop

One strategy that fungi and bacteria use to degrade wood is called tunneling [75]. During this process, fungal hyphae or bacteria drill parallel to the secondary lamella (cell walls) of the wood fibers. Behind the microbial woodworkers are empty tunnels with walls consisting of biofilm and partially degraded lignin [75]. Invasion of the tunnels by other bacteria then proceeds.

This tunneling process is mechanically depicted in Figure G-1. Figure G-1 shows a fungal hyphae which emits extracellular enzymes through a biofilm to degrade the material in front of it and grow via exocitosis. After taking up dissolved nutrients, the hyphae grows into the void by using actin, dynein and kinase motors in the spitzenkoerper [80] to transport exocitic and endocitic vesicles to the hyphae tip. These vesicles secrete enzymes which cleave crosslinks in the wall, reducing the yield stress of the hyphae tip. Stress in the cell wall from internal cytoplasmic turgor pressure then plastically deforms the tip [167]. Simultaneously, the exocitic enzymes and vesicles add new wall material to the hypha tip. The process repeats when the fungus can move forward again.

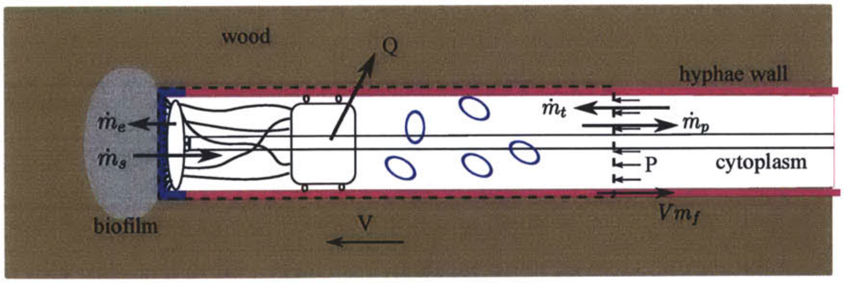

Figure G-1: Control volume model of a fungal hypha. P: internal turgor pressure. m_e mass of extracellular enzymes. m_t mass of nutrients transported to hypha. m_p mass of product. Vm_f flux of cell wall material advected out of the control volume. m_s: mass of dissolved substrate uptake. V: speed. Q rate of energy loss due to inefficiency and maintenance.

The mechanics of this process is rather straightforwardly modeled using a control volume approach and I use it to come up with a functional form for the speed V at which the fungus decays wood. I assume a cylindrical fungus with a cross sectional area A that moves forward and erodes material at the velocity V. There is an internal turgor pressure P which pushes the tip forward. The fungus is comprised of an array of different organic compounds having a total mass in the control volume m_f, with an average chemical potential g_f. The total mass of incoming substrate is $m_s = A\rho_s V$ where ρ_s is the substrate density. The average chemical potential of the substrate is g_s. The mass of reaction products which exit the control volume is m_p and their average chemical potential is g_p. Nutrients and other compounds m_t may be transported to the hyphae from other parts of the fungus. I draw the control volume large enough to contain all of the vacuoles and active machinery within the hyphae so internal motor driven transport processes need not be modeled. Within the control volume (dashed lines) of Figure G-1, mass conservation is

$$\dot{m}_f = \dot{m}_s + \dot{m}_t - \dot{m}_p - \dot{m}_e, \tag{G.1}$$

where m_e is mass of extracellular enzymes emitted to the biofilm.

The first law of thermodynamics applied to the control volume is

$$\dot{m}_f g_f + m_f \dot{g_f} = A\rho_s V g_s + \dot{m}_t g_t - \dot{m}_p g_p - APV - \dot{m}_e g_e - Q - A_w \rho_f V g_f. \tag{G.2}$$

The first two terms represent the change internal system energy through changes in mass and chemical energy, either by biosynthesis or ATP conversion. The first term on the right represents incoming chemical energy, the second term represents the transport of mass and energy to the cell from other cells in the hyphae network, the third term represents the flux of products out of the system, the fourth term represents the work done to elongate the cell wall (mobility work), the fifth term represents the energy lost from the system via release of extracellular enzymes, the sixth term represents heat flux out of th system due to inefficiencies and maintenance, and the final term represents advection of chemical energy out of the control volume. This final term is also related to the motility work, as mass is added to the wall (and later advected out) in order to move. A_w is the cross sectional area of the fungal wall as I assume the spitzenkoerper, proteins, DNA, and other active cellular components flow with the hyphae tip [80] and do not advect out of the control volume. P is the internal cell turgor pressure.

The relationship between speed V and enzyme release $\dot{m}_e$ can be modeled via diffusion of enzyme and dissolved substrate through the biofilm, but that is not necessary here.

A few assumptions can be made to proceed with the analysis. Assuming that no enzymes or substrate are lost to advection in the annulus between the fungus and cell wall, and that extracellular enzymes and proteins are recycled by the fungus [80], we can set that enzyme flux to zero

$$\dot{m}_e = 0. \tag{G.3}$$

Assuming steady state,

$$\dot{m}_f = 0, \quad \dot{g}_f = 0 \tag{G.4}$$

or steady state growth,

$$\dot{m}_f = b, \dot{g}_f = 0, \tag{G.5}$$

leads to additional simplifications. We can also assume that the hypha is self sustaining and nutrients are not being transported from other hyphae cells,

$$\dot{m}_t = 0, \tag{G.6}$$

although mass and energy may transport out as product $\dot{m}_p g_p$.

Assuming that the hyphae is self sustaining, in steady state and recycling extracellular enzymes, inserting (G.6), (G.3), and (G.4) into (G.1) results in

$$\dot{m}_s = \dot{m}_p. \tag{G.7}$$

Inserting (G.7), (G.4), and (G.3) into (G.2) reduces the system to

$$0 = A\rho_s V \Delta G - AFV - Q - A_w \rho_f V g_f \tag{G.8}$$

ΔG is the average change in free energy between substrate and microbial products. The first term represents incoming energy to the system, the second and fourth terms are the work due to microbial mobility, and the third term Q, is associated with the efficiency of processes and maintenance. Steady growth can be modeled by adding the term bg_f on the left hand side and including b in equation (G.7),

$$bg_f = A\rho_s V \Delta G + bg_p - APV - Q - A_w \rho_f V g_f \tag{G.9}$$

Empirical observations find that the yield stress of a polymers scale linearly with the strain rate [168], but with a very shallow slope [169]. This indicates the yield stress is rather constraint over a wide range of strain rates. Furthermore, plastic deformation of the cell wall occurs discretely rather than as a continuous process [56,80]. Both behaviors suggest modeling the yield strength required to plastically deform the new cell wall material as a constant, indicating that the fungal cell likely does not need to increase turgor pressure to increase growth speed V. The yield stress however may indirectly be a function of V since it depends on the rate of exocitic and endocitic vesicle transport [80].

Therefore we can express the velocity as a function of Q, F, $\Delta G, (\Delta G = g_s - g_p \approx g_f - g_p)$, and microbial geometry.

$$V = \frac{Q + b\Delta G}{A\rho_s \Delta G - (AP + A_w \rho_f g_f)} \tag{G.10}$$

Equation (G.10) shows that under steady state growth, the velocity V results from a balance between energy required for maintenance and growth, and the remaining energy from the substrate after subtracting the work required for transportation. While this is a nice way to show energy balance in cells, it actually tells us nothing about what controls the speed V. Inclusion of unsteady terms provide no additional information in this regard. Rather, the speed at which the fungus moves is determined by the release of extracellular enzymes $\dot{m}_e$, which is determined by the signals and switches in the pathways that control cellular growth $\dot{m}_f$, energy storage $\dot{g}_f$, and waste $\dot{m}_p g_p$, as cells may process substrate rapidly but not fully respire or utilize them, dumping them out with high chemical potential.

As usual mechanics will not tell us the kinetics of biodegradation, unless there is perhaps a consistent rate-limiting physical process. This is a subtle but rather obvious statement. Similarly, thermodynamics can tell you how much gas I will use when I drive my 2.0L turbocharged WRX downhill into a headwind at a constant velocity. But the laws of thermodynamics cannot tell you that I wanted a speeding ticket for driving 100 mph. The speed is controlled by the various inputs, signals, chemical state of my brain, and conditions of my surrounding environment.

Similarly, the rates at which decomposition proceeds are related to the inputs and signals to the community of decomposers and the physical processes of various length and timescales which shape their environment. The work presented in this thesis scratches the surface of the 'systems ecology' of natural ecosystems.

Bibliography

[1] Lutzow, et al. (2006) Stabilization of organic matter in temperate soils: mechanisms and their relevance under different soil conditions - a review. *European Journal of Soil Science* 57:426–445.

[2] Cox PM (2001) Description of the triffid dynamics global vegetation model, tech. note 24., (Hadley Centre, Met Office), Technical report.

[3] Moorcroft PR, Hurtt GC, Pacala SW (2001) A method for scaling vegetation dynamics: The ecosystem demography model (ed). *Ecological Monographs* 71:557–585.

[4] Medvigy D, Wofsy SC, Munger JW, Hollinger DY, Moorcroft PR (2009) Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem demography model version 2. *Journal of Geophysical Research* 114:G01002+.

[5] Sitch S, et al. (2003) Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the lpj dynamic global vegetation model. *Global Change Biology* 9:161–185.

[6] Krinner G (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles* 19:GB1015+.

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

[8] Manzoni S, Trofymow JA, Jackson RB, Porporato A (2010) Stoichiometric controls on carbon, nitrogen, and phosphorus dynamics in decomposing litter. *Ecological Monographs* 80:89–106.

[9] Manzoni S, Porporato A (2009) Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biology and Biochemistry* 41:1355–1379.

[10] Agren GI, Bosatta E (1998) *Theoretical Ecosystem Ecology: Understanding Element Cycles* (Cambridge University Press), 1 edition.

[11] Jastrow, et al. (2007) Mechanisms controlling soil carbon turnover and their potential application for enhancing carbon sequestration. *Climatic Change* 80:5–23.

[12] Prescott CE (2010) Litter decomposition: what controls it and how can we alter it to sequester more carbon in forest soils? *Biogeochemistry* 101:133–149.

[13] Forney DC, Rothman DH (2012) Common structure in the heterogeneity of plant matter decay. *Journal of the Royal Society Interface* pp In Press.+.

[14] Carpenter S (1981) Decay of heterogenous detritus: A general model. *Journal of Theoretical Biology* 89:539–547.

[15] Bosatta E (1985) Theoretical analysis of decomposition of heterogeneous substrates. *Soil Biology and Biochemistry* 17:601–610.

[16] Tarutis WJ (1994) A mean-variance approach for describing organic matter decomposition. *Journal of Theoretical Biology* 168:13–18.

[17] Janssen BH (1984) A simple method for calculating decomposition and accumulation of 'young' soil organic matter. *Plant and Soil* 76:297–304.

[18] Bolker BM, Pacala SW, Parton WJ (1998) Linear analysis of soil decomposition: Insights from the century model. *Ecological Applications* 8:425–439.

[19] Harmon M (2007) *http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=TD023.*

[20] Harmon M, Lidet (1995) Meeting the challenges of long-term, broad-scale ecological experiments. *LTER Network Office* 19.

[21] Gholz HL, Wedin DA, Smitherman SM, Harmon ME, Parton WJ (2000) Long-term dynamics of pine and hardwood litter in contrasting environments: toward a global model of decomposition. *Global Change Biology* 6:751–765.

[22] Adair EC, et al. (2008) Simple three-pool model accurately describes patterns of long-term litter decomposition in diverse climates. *Global Change Biology* 14:2636–2660.

[23] Cusack DF, Chou WW, Yang WH, Harmon ME, Silver WL (2009) Controls on long-term root and leaf litter decomposition in neotropical forests. *Global Change Biology* 15:1339–1355.

[24] Harmon ME, et al. (2009) Long-term patterns of mass loss during the decomposition of leaf and fine root litter: an intersite comparison. *Global Change Biology* 15:1320–1338.

[25] Currie WS, et al. (2010) Cross-biome transplants of plant litter show decomposition models extend to a broader climatic range but lose predictability at the decadal time scale. *Global Change Biology* 16:1744–1761.

[26] Lloyd J, Taylor JA (1994) On the temperature dependence of soil respiration. *Functional Ecology* 8:315–323.

[27] Brown JH, Gillooly JF, Allen AP, Savage VM, West GB (2004) Toward a metabolic theory of ecology. *Ecology* 85:1771–1789.

[28] Ågren G, Bosatta E (2002) Reconciling differences in predictions of temperature response of soil organic matter. *Soil Biology and Biochemistry* 34:129–132.

[29] Craine JM, Fierer N, McLauchlan KK (2010) Widespread coupling between the rate and temperature sensitivity of organic matter decay. *Nature Geoscience* 3:854–857.

[30] Liski J, Ilvesniemi H, Mäkelä A, Westman CJ (1999) Co emissions from soil in response to climatic warming are overestimated: The decomposition of old soil organic matter is tolerant of temperature. *Ambio* 28:171–174.

[31] Kirschbaum MUF (2009) The temperature dependence of organic matter decomposition: seasonal temperature variations turn a sharp short-term temperature response into a more moderate annually averaged response. *Global Change Biology* 9999.

[32] Berg B, Laskowski R (2006) *Litter decomposition : a guide to carbon and nutrient turnover* (Elsevier, Amsterdam ; London) Vol. 38.

[33] Manzoni S, Jackson RB, Trofymow JA, Porporato A (2008) The global stoichiometry of litter nitrogen mineralization. *Science* 321:684–686.

[34] Minderman G (1968) Addition, decomposition and accumulation of organic matter in forests. *The Journal of Ecology* 56:355–362.

[35] Melillo JM, Aber JD, Muratore JF (1982) Nitrogen and lignin control of hardwood leaf litter decomposition dynamics. *Ecology* 63:621–626.

[36] Yang HS, Janssen BH (2000) A mono-component model of carbon mineralization with a dynamic rate constant. *European Journal of Soil Science* 51:517–529.

[37] Trumbore S (2000) Age of soil organic matter and soil respiration: Radiocarbon constraints on belowground c dynamics. *Ecological Applications* 10:399–411.

[38] Melillo J, et al. (1989) Carbon and nitrogen dynamics along the decay continuum: Plant litter to soil organic matter. *Plant and Soil* 115:189–198.

[39] Manzoni S, Katul GG, Porporato A (2009) Analysis of soil carbon transit times and age distributions using network theories. *Journal of Geophysical Research* 114:G04025+.

[40] Feng Y (2009) Fundamental considerations of soil organic carbon dynamics. *Soil Science* 174:467–481.

[41] Feng Y (2009) K-model-a continuous model of soil organic carbon dynamics: Theory. *Soil Science* 174:482–493.

[42] Berner RA (1980) *Early Diagenesis* (Princeton University Press).

[43] Jenkinson DS (1977) Studies on the decomposition of plant material in soil. v. the effect of plant cover and soil type on the loss of carbon from 14c labeled ryegrass decomposing under field conditions. *Journal of Soil Science* 28:424–434.

[44] van Veen JA, Paul EA (1981) Organic carbon dynamics in grassland soils. 1. background information and computer simulation. *Canadian Journal of Soil Science* 61:185–201.

[45] Parton WJ, Schimel DS, Cole CV, Ojima DS (1987) Analysis of factors controlling soil organic matter levels in great plains grasslands. *Soil Sci Soc Am J* 51:1173–1179.

[46] Parton WJ, et al. (1993) Observations and modeling of biomass and soil organic matter dynamics for the grassland biome worldwide. *Global Biogeochemical Cycles* 7:785–809.

[47] Eijsackers H, Zehnder A (1990) Litter decomposition: a russian matriochka doll. *Biogeochemistry* 11:153–174.

[48] Beare MH, et al. (1992) Microbial and faunal interactions and effects on litter nitrogen and decomposition in agroecosystems. *Ecological Monographs* 62.

[49] Boudreau BP, Ruddick BR (1991) On a reactive continuum representation of organic matter diagenesis. *Am J Sci* 291:507–538.

[50] Rothman DH, Forney DC (2007) Physical model for the decay and preservation of marine organic carbon. *Science* 316:1325–1328.

[51] Forney D, Rothman D (2012) Inverse method for estimating respiration rates from decay time series. *Biogeosciences Discussion* 9:Published by Mar 28, 2012.

[52] Forney D, Rothman D (2012) Residence of organic carbon in degradation networks. *TBD* p In prep.

[53] Wackett L (2006) in *The Prokaryotes*, eds Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (Springer New York), pp 956–968.

[54] Denman KL, et al. (2007) in *Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, eds Solomon S, et al. (Cambridge University Press, Cambridge UK/New York), pp 499–587.

[55] Benner R, Herndl GJ (2011) Bacterially derived dissolved organic matter in the microbial carbon pump.

[56] Sheetz M (2011) Course notes. *Cell as a Machine, MIT course 2.S997.*

[57] Trevors JT (2010) One gram of soil: a microbial biochemical gene library. *Antonie van Leeuwenhoek* 97:99–106.

[58] Horner-Devine MC, Carney KM, Bohannan BJM (2004) An ecological perspective on bacterial biodiversity. *Proc. R. Soc. Lond. B* 271:113–122.

[59] Curtis TP, Sloan WT, Scannell JW (2002) Estimating prokaryotic diversity and its limits. *Proceedings of the National Academy of Sciences* 99:10494–10499.

[60] Claire Horner-Devine M, Leibold MA, Smith VH, Bohannan BJM (2003) Bacterial diversity patterns along a gradient of primary productivity. *Ecology Letters* 6:613–622.

[61] Schlesinger WH, Andrews JA (2000) Soil respiration and the global carbon cycle. *Biogeochemistry* 48:7–20.

[62] Madigan MT, Martinko JM, Dunlap PV, Clark DP (2005) *Brock Biology Of Microorganisms* (Benjamin Cummings), 11th edition.

[63] Tenney FG, Waksman SA (1929) Composition of natural organic materials and their decomposition in the soil: Iv. the nature and rapidity of decomposition of the various organic complexes in different plant materials, under aerobic conditions. *Soil Science* 28:55–84.

[64] Burdige DJ (2006) *Geochemistry of Marine Sediments* (Princeton University Press).

[65] Berg B, McClaugherty C (2007) *Plant Litter: Decomposition, Humus Formation, Carbon Sequestration* (Springer), 2nd edition.

[66] Lee C, Arnosti C, Wakeham S (2004) Particulate organic matter in the sea: The composition conundrum. *Ambio* 33:565–575.

[67] Paul EA (2007) *Soil Microbiology, Ecology and Biochemistry, Third Edition* (Academic Press), 3 edition.

[68] Andrén O, Paustian K (1987) Barley straw decomposition in the field: A comparison of models. *Ecology* 68.

[69] Oades J (1988) The retention of organic matter in soils. *Biogeochemistry* 5:35–70.

[70] Hedges JI, Oades JM (1997) Comparative organic geochemistries of soils and marine sediments. *Organic Geochemistry* 27:319–361.

[71] Mayer L (1994) Relationships between mineral surfaces and organic carbon concentrations in soils and sediments. *Chemical Geology* 114:347–363.

[72] Vetter YA, Deming JW, Jumars PA, Krieger-Brockett BB (1998) A predictive model of bacterial foraging by means of freely released extracellular enzymes. *Microbial Ecology* 36:75–92.

[73] Nieder R, Benbi DK (2008) *Carbon and Nitrogen in the Terrestrial Environment* (Springer), 1 edition.

[74] Semple K, et al. (2004) Defining bioavailability and bioaccessibility of contaminated soil and sediment is complicated. *Environmental Science & Technology* 38:228A–231A.

[75] Blanchette RA (1995) Degradation of the lignocellulose complex in wood. *Canadian Journal of Botany* 73:999–1010.

[76] Hättenschwiler S, Tiunov AV, Scheu S (2005) Biodiversity and litter decomposition in terrestrial ecosystems. *Annual Review of Ecology, Evolution, and Systematics* 36.

[77] Blair JM, Parmelee RW, Beare MH (1990) Decay rates, nitrogen fluxes, and decomposer communiies of single- and mixed-species foliar litter. *Ecology* 71:1976–1985.

[78] Van Der Heijden MGA, Bardgett RD, Van Straalen NM (2008) The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters* 11:296–310.

[79] Allison SD (2005) Cheaters, diffusion and nutrients constrain decomposition by microbial enzymes in spatially structured environments. *Ecology Letters* 8:626–635.

[80] Steinberg G (2007) Hyphal growth: a tale of motors, lipids, and the spitzenkorper. *Eukaryotic Cell* 6:351–360.

[81] Riquelme M, Reynaga-Peñaa CG, Gierzb G, Bartnicki-García S (1998) What determines growth direction in fungal hyphae? *Fungal Genetics and Biology* 24:101–109.

[82] Bosatta E, Agren GI (1996) Theoretical analyses of carbon and nutrient dynamics in soil profiles. *Soil Biology and Biochemistry* 28:1523–1531.

[83] van Keulen H (2001) (tropical) soil organic matter modelling: problems and prospects. *Nutrient Cycling in Agroecosystems* 61:33–39.

[84] Loreau M (2001) Microbial diversity, producerdecomposer interactions and ecosystem processes: a theoretical model. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 268:303–309.

[85] Schimel JP, Weintraub MN (2003) The implications of exoenzyme activity on microbial carbon and nitrogen limitation in soil: a theoretical model. *Soil Biology and Biochemistry* 35:549–563.

[86] Ash RB (2008) *Basic Probability Theory (Dover Books on Mathematics)* (Dover Publications).

[87] Bharucha-Reid AT (2010) *Elements of the Theory of Markov Processes and Their Applications (Dover Books on Mathematics)* (Dover Publications).

[88] Matis JH, Wehrly TE (1998) A general approach to non-markovian compartmental models. *Journal of Pharmacokinetics and Pharmacodynamics* 26:437–456.

[89] Matis JH, Wehrly TE (1990) Generalized stochastic compartmental models with erlang transit times. *Journal of Pharmacokinetics and Pharmacodynamics* 18:589–607.

[90] Faddy MJ (1990) Compartmental models with phase-type residence-time distributions. *Appl. Stochastic Models Data Anal.* 6:121–127.

[91] Neuts MF (1995) *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach* (Dover Publications), Revised edition.

[92] van Veen J, Ladd J, Frissel M (1984) Modelling c and n turnover through the microbial biomass in soil. *Plant and Soil* 76:257–274.

[93] Jenkinson DS, Andrew SPS, Lynch JM, Goss MJ, Tinker PB (1990) The turnover of organic carbon and nitrogen in soil. *Philosophical Transactions: Biological Sciences* 329:361–368.

[94] Liang C, Cheng G, Wixon D, Balser T (2010) An absorbing markov chain approach to understanding the microbial role in soil carbon stabilization. *Biogeochemistry* pp 1–7.

[95] Friedlingstein P, et al. (2006) Climatecarbon cycle feedback analysis: Results from the C4MIP model intercomparison. *Journal of Climate* 19:3337–3353.

[96] Todd-Brown K, Hopkins F, Kivlin S, Talbot J, Allison S (2011) A framework for representing microbial decomposition in coupled climate models. *Biogeochemistry* pp 1–15.

[97] Arneth A, Mercado L, Kattge J, Booth B (2012) Future challenges of representing land-processes in studies on land-atmosphere interactions. *Biogeosciences Discussions* 9:3545–3577.

[98] Luo Y, et al. (2003) Sustainability of terrestrial carbon sequestration: A case study in duke forest with inversion approach. *Global Biogeochemical Cycles* 17:1021+.

[99] Harmon M (2007) LTER Intersite Fine Litter Decomposition Experiment (LIDET), Forest Science Data Bank code TD023, Corvallis, OR. 2007, [Database] http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=TD023.

[100] Falkowski P, et al. (2000) The global carbon cycle: A test of our knowledge of Earth as a system. *Science* 290:291–296.

[101] Trumbore S (2009) Radiocarbon and soil carbon dynamics. *Annual Review of Earth and Planetary Sciences* 37:47–66.

[102] Johnson DW, Miller WW, Rau BM, Meadows MW (2011) The nature and potential causes of nutrient hotspots in a sierra nevada forest soil. *Soil Science* 176:596–610.

[103] Middelburg JJ (1989) A simple rate model for organic matter decomposition in marine sediments. *Geochim. Cosmochim. Acta* 53:1577–1581.

[104] Kirschbaum MUF (2000) Will changes in soil organic carbon act as a positive or negative feedback on global warming? *Biogeochemistry* 48:21–51.

[105] May RM (1975) in *Ecology and Evolution of Communities*, eds Cody ML, Diamond JM (Harvard University Press, Cambridge, MA), pp 81–120.

[106] Vlad MO, Huber DL, Ross J (1997) Rate statistics and thermodynamic analogies for relaxation processes in systems with static disorder: Application to stretched exponential. *The Journal of Chemical Physics* 106:4157–4167.

[107] Ross J, Vlad MO (1999) Nonlinear kinetics and new approaches to complex reaction mechanisms. *Annual Review of Physical Chemistry* 50:51–78.

[108] Plonka A (2001) *Dispersive Kinetics* (Springer), 1 edition.

[109] Bosatta E (1995) The power and reactive continuum models as particular cases of the q-theory of organic matter dynamics. *Geochimica et Cosmochimica Acta* 59:3833–3835.

[110] Gessner MO, et al. (2010) Diversity meets decomposition. *Trends in Ecology & Evolution* 25:372–380.

[111] Feller W (1971) *An Introduction to Probability Theory and Its Applications, Vol. 2* (Wiley), 2nd edition.

[112] Lamanna R (2005) On the inversion of multicomponent NMR relaxation and diffusion decays in heterogeneous systems. *Concepts Magn. Reson.* 26A:78–90.

[113] Kroeker RM, Henkelman RM (1986) Analysis of biological NMR relaxation data with continuous distributions of relaxation times. *Journal of Magnetic Resonance (1969)* 69:218–235.

[114] Ansari A, et al. (1985) Protein states and proteinquakes. *Proceedings of the National Academy of Sciences* 82:5000–5004.

[115] Hansen PC (1987) *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion (Monographs on Mathematical Modeling and Computation)* (Society for Industrial Mathematics).

[116] Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1992) *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge University Press), 2 edition.

[117] Gallegos DP, Smith DM (1988) A NMR technique for the analysis of pore structure: Determination of continuous pore size distributions. *Journal of Colloid and Interface Science* 122:143–153.

[118] Kleinberg R (1996) Utility of NMR T2 distributions, connection with capillary pressure, clay effect, and determination of the surface relaxivity parameter ρ_2. *Magnetic Resonance Imaging* 14:761–767.

[119] Limpert E, Stahel WA, Abbt M (2001) Log-normal distributions across the sciences: Keys and clues. *BioScience* 51:341–352.

[120] Montroll EW, Shlesinger MF (1982) On 1/f noise and other distributions with long tails. *PNAS* 79:3380–3383.

[121] Yeramian E, Claverie P (1987) Analysis of multiexponential functions without a hypothesis as to the number of components. *Nature* 326:169–174.

[122] Bolin B, Rodhe H (1973) A note on the concepts of age distribution and transit time in natural reservoirs. *Tellus* 25:58–62.

[123] Kendall M, Gibbons JD (1990) *Rank Correlation Methods* (Oxford University Press, New York), 5 edition.

[124] Hansen PC (1994) Regularization tools: A Matlab package for analysis and solution of discrete ill-posed problems. *Numerical Algorithms* 6:1–35.

[125] Rodhe H (1992) in *Global Biogeochemical Cycles*, eds Butcher SS, Charlson RJ, Orians GH, Wolfe GV (Academic Press, San Diego), pp 55–72.

[126] Olson JS (1963) Energy storage and the balance of producers and decomposers in ecological systems. *Ecology* 44:322–331.

[127] Feng Y (2009) K-model-a continuous model of soil organic carbon dynamics: Model parameterization and testing. *Soil Science* 174:494–507.

[128] Strang G (1986) *Introduction to Applied Mathematics* (Wellesley-Cambridge Press), First edition, later impression edition.

[129] Mathworks (2009) mldivide.m. *Matlab Help.*

[130] Mathworks (2011) Matlab help: pinv.m. *Matlab Help.*

[131] Konishi S, Kitagawa G (2010) *Information Criteria and Statistical Modeling (Springer Series in Statistics)* (Springer), Softcover reprint of hardcover 1st ed. 2008 edition.

[132] Zhou Y, Zhuang X (2006) Robust reconstruction of the rate constant distribution using the phase function method. *Biophysical Journal* 91:4045–4053.

[133] Istratov AA, Vyvenko OF (1999) Exponential analysis in physical phenomena. *Review of Scientific Instruments* 70:1233–1257.

[134] Pierce JR (1980) *An Introduction to Information Theory: Symbols, Signals and Noise* (Dover Publications), 2nd revised ed. edition.

[135] Faddy MJ (1993) A structured compartmental model for drug kinetics. *Biometrics* 49:243–248.

[136] Kendall DG (1953) Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *Annals of Mathematical Statistics* 24:338–354.

[137] Forrester JW (1961) *Industrial Dynamics* (The M.I.T. Press and John Wiley & Sons), 1st edition.

[138] Godfrey K (1983) *Compartmental Models and Their Application* (Academic Press, New York).

[139] Matis JH, Wehrly TE, Ellis WC (1989) Some generalized stochastic compartment models for digesta flow. *Biometrics* 45.

[140] Carr SA, Baird RB (2000) Mineralization as a mechanism for toc removal: study of ozone/ozoneperoxide oxidation using ft-ir. *Water Research* 34:4036–4048.

[141] Matis JH, Wehrly TE, Metzler CM (1983) On some stochastic formulations and related statistical moments of pharmacokinetic models. *Journal of Pharmacokinetics and Pharmacodynamics* 11:77–92.

[142] Greenberg M (1998) *Advanced Engineering Mathematics (2nd Edition)* (Prentice Hall), 2 edition.

[143] Franklin G, Powell JD, Emami-Naeini A (2005) *Feedback Control of Dynamic Systems (5th Edition)* (Prentice Hall), 5 edition.

[144] Bosatta E, Agren GI (1991) Dynamics of carbon and nitrogen in the organic matter of the soil: A generic theory. *The American Naturalist* 138:227–245.

[145] Zhou Y, Zhuang X (2007) Kinetic analysis of sequential multistep reactions. *Journal of Physical Chemistry B* 111:13600–13610.

[146] Cheng W (2009) Rhizosphere priming effect: Its functional relationships with microbial turnover, evapotranspiration, and cn budgets. *Soil Biology and Biochemistry* 41:1795–1801.

[147] Smith OL (1979) An analytical model of the decomposition of soil organic matter. *Soil Biology and Biochemistry* 11:585–606.

[148] Efron B, Tibshirani RJ (1994) *An Introduction to the Bootstrap (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)* (Chapman and Hall/CRC), 1 edition.

[149] Research R (2006) *Rothamstead Research: Guide to the Classical and other Long-term Experiments, Datasets and Sample Archive* (Lawes Agricultural Trust Co. Ltd).

[150] Leipnik RB (1991) On lognormal random variables: I-the characteristic function. *Journal of the Australian Mathematical Society Series B* 32:327–347.

[151] Beck C, Cohen E (2003) Superstatistics. *Physica A: Statistical Mechanics and its Applications* 322:267–275.

[152] Nadarajah S (2008) Comment on a general model for long-tailed network traffic approximation. *The Journal of Supercomputing* 44:98–101.

[153] Holgate P (1989) The lognormal characteristic function. *Communications in Statistics - Theory and Methods* 18:4539–4548.

[154] Mullen RE (1998) The lognormal distribution of software failure rates: application to software reliability growth modeling. *Software Reliability Engineering, 1998. Proceedings. The Ninth International Symposium on* pp 134–142.

[155] Montroll EW, Bendler JT (1984) On lévy (or stable) distributions and the williams-watts model of dielectric relaxation. *Journal of Statistical Physics* 34:129–162.

[156] Allella F, Chiodo E, Lauria D, Pagano M (2001) Negative log-gamma distribution for data uncertainty modelling in reliability analysis of complex systems - methodology and robustness. *International Journal of Quality & Reliability Management* pp 307–323.

[157] Bulmer MG (1974) On fitting the poisson lognormal distribution to species-abundance data. *Biometrics* 30:101–110.

[158] Grundy PM (1951) The expected frequencies in a sample of an animal population in which the abundances of species are log-normally distributed. part i. *Biometrika* 38:427–434.

[159] Rossberg AG (2008) Laplace transforms of probability distributions and their inversions are easy on logarithmic scales. *Journal of Applied Probability* 45:531–541.

[160] Merico A, Bruggeman J, Wirtz K (2009) A trait-based approach for downscaling complexity in plankton ecosystem models. *Ecological Modelling* 220:3001–3010.

[161] Bruggeman J (2010) Decay of a continuum of organic matter types, characterized by decay rate. *Email Correspondance.*

[162] Norberg J, et al. (2001) Phenotypic diversity and ecosystem functioning in changing environments: A theoretical framework. *Proceedings of the National Academy of Sciences of the United States of America* 98:11376–11381.

[163] Wirtz KW, Eckhardt B (1996) Effective variables in ecosystem models with an application to phytoplankton succession. *Ecological Modelling* 92:33–53.

[164] Crow EL, Shimizu K (1987) *Lognormal Distributions (Statistics: A Series of Textbooks and Monographs)* (CRC Press), 1 edition.

[165] Viberg M (1995) Subspace-based methods for the identification of linear time-invariant systems. *Automatica* 31:1835–1851.

[166] Sarkar TK, Pereira O (1995) Using the matrix pencil method to estimate the parameters of a sum of complex exponentials. *Antennas and Propagation Magazine, IEEE* 37:48–55.

[167] Hasek J, Geitmann A (2000) *Cell Biology of Plant and Fungal Tip Growth (Nato: Life and Behavioural Sciences, 328)* (Ios Pr Inc).

[168] Argon AS (1973) A theory for the low-temperature plastic deformation of glassy polymers. *Philosophical Magazine* 28:839–865.

[169] Boyce MC, Parks DM, Argon AS (1988) Large inelastic deformation of glassy polymers. part i: rate dependent constitutive model. *Mechanics of Materials* 7:15–33.