https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

Biogeosciences Discussions
Open Access
EGU

Theoretical insights from upscaling Michaelis-Menten microbial dynamics in biogeochemical models: a dimensionless approach

Chris H. Wilson[1], Stefan Gerber[2]

[1]Agronomy Department, University of Florida, Gainesville, FL, USA 32611
[2]Soil and Water Sciences Department, University of Florida, Gainesville, FL, USA 32611

Correspondence to: Chris H. Wilson (chwilson@ufl.edu)

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

Abstract. Leading an effective response to the accelerating crisis of anthropogenic climate change will require improved understanding of global carbon cycling. A critical source of uncertainty in Earth Systems Models (ESMs) is the role of microbes in mediating both the formation and decomposition of soil organic matter, and hence in determining patterns of CO2 efflux. Traditionally, ESMs model carbon turnover as a first order process impacted primarily by abiotic factors, whereas contemporary biogeochemical models often explicitly represent the microbial biomass and enzyme pools as the active agents of decomposition. However, the combination of non-linear microbial kinetics and ecological heterogeneity across space guarantees that upscaled dyamics will violate mean-field assumptions via Jensen's Inequality. Violations of mean-field assumptions mean that parameter estimates from models fit to upscaled data (e.g. eddy covariance towers) are likely systematically biased. Here we present a generic mathematical analysis of upscaling michaelis-menten kinetics under heterogeneity, and provide solutions in dimensionless form. We illustrate how out dimensionless form facilitates qualitative insight into the significance of this scale transition, and argue that it will facilitate cross site intercomparisons of flux data. We also identify the critical terms that need to be constrained in order to unbias parameter estimates.

1 Introduction

The current crisis of anthropogenic climate change is expected to accelerate during the 21st century. Despite considerable effort to better constrain global biogeochemical models, considerable uncertainty remains about how best to represent emerging mechanistic understanding of soil element cycling into process-based models (Wieder et al. 2015, 2018; Todd-Brown, Zheng, and Crowther 2018). This is a critical gap in knowledge because variations among models predict hugely varying responses to global change drivers such as temperature, soil moisture, and CO_2 enrichment. For example, a traditional first-order linear model forecasts no change or even slight enhancement of soil organic carbon (SOC) pools by 2100 whereas one microbial-explicit model forecasts a loss of ~70Pg of carbon (C), depending on whether microbial physiology acclimates to higher temperatures (Wieder, Bonan, and Allison 2013). In general, our understanding of how carbon (and other elements) cycles in soil is undergoing significant revision toward a more microbial-centric paradigm. In contrast to traditional first-order linear models (e.g. CENTURY, Parton et al. 1987), microbial explicit models feature non-linear dynamics in which microbial biomass (or, similarly, microbially-driven enzyme pools) are responsible for decomposition, in addition to providing substrate for synthesis of potentially long-term SOC (Blankinship and Schimel 2018; Blankinship et al. 2018). While indisputably a better representation of our scientific knowledge, non-linear microbial models face several well-known challenges, including less analytical tractability, greater computational challenges, and uncertainty about structural formulation and dynamics (Georgiou et al. 2017; Sihi et al. 2016). However, one critical consequence of non-linear microbial models that is only recently gaining attention is their implications for addressing the upscaling challenge.

While the fields of population and community ecology have long confronted the challenges posed by non-linearity and heterogeneity in spatiotemporal scaling of ecological dynamics (Chesson 2009; Levin 1992), ecosystem ecology and biogeochemistry have tended to approach the challenge of scale either by 1) utilizing mean-field assumptions, or 2) addressing the challenge of scaling via grid-based computational/numeric methods. While there is nothing

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

wrong inherently with either approach, they unfortunately cannot yield theoretical insight into the consequences of non-linearity and heterogeneity for scaling. Briefly, the combination of non-linearity and heterogeneity means that aggregated behavior differs systematically from mean-field predictions, a special case of Jensen's Inequality. In mathematical notation:

$$\mathrm{E}[f(x)] \neq f(\mathrm{E}[x]) \tag{1}$$

Although Jensen's Inequality is well-known from basic probability theory (Ross 2002), it's implications for ecological dynamics under heterogeneity were not well-appreciated until the pioneering work of Peter Chesson in the 1990s (Chesson 1998). In the case of carbon cycle science, there are a few immediate and critical applications. For instance, most trace gas emission processes are well-known to be non-linear functions of underlying drivers such as temperature and soil moisture. For example, ecosystem respiration (hereafter "R_{eco}") is an exponential function of temperature (usually expressed in Q_{10}), and a unimodal function of soil moisture. Thus, when matching observations of CO_2 efflux ("F") to ecosystems, variations in soil temperature and moisture could imply that F differs systematically from a mean-field prediction. In addition to missing critical analytical insight, not accounting for this behavior might have severe consequences for inverse modeling and estimation of the parameters governing process-based models (PBMs). The basic consequences of Jensen's Inequality for estimation of trace gas emission (CH_4 and N_2O) were first discussed by Van Oijen et al. (2017), but have not been picked up on elsewhere, until the present work and by Chakrawal et al. (2019).

Chakrawal et al. (2019) provide a detailed and compelling first-pass application of scale transition theory to biogeochemical modeling. Our contribution here complements their laudable effort by providing a more generic mathematical analysis of the scale transition, equally applicable to both forward and reverse michaelis-menten microbial kinetics. As in Chakrawal et al. (2019), we address the consequences of heterogeneity in both substrate/microbes ("biochemical heterogeneity") as well as in the kinetic parameters ("ecological heterogeneity"). However, we diverge from their approach in that, rather than explore detailed simulation models, we derive a completely non-dimensionalized expression for aggregating non-linear microbial kinetics over both types of heterogeneity simultaneously. We illustrate the enormous clarity this brings in several special cases of our full analysis. Altogether, our approach provides universal insight into the properties of the scale transition, and enables clear conclusions to be drawn across systems in terms of the role of spatial variances and covariances in shaping ecosystem carbon efflux. Our work provides a simplified, yet systematic framework around which to base subsequent empirical and simulation-based studies.

2. Carbon Efflux and the Scale Transition

A variety of microbial-explicit PBMs have been proposed in the literature, starting with the classic enzyme pool model of Schimel and Weintraub (2003). In order to elucidate universal properties of the scale transition, we focus here on the CO_2 efflux following decomposition of a single substrate by a single microbial pool obeying Michaelis-Menten (MM) dynamics:

$$F = -f(C, MB, \theta) \tag{2}$$

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

where F is the CO_2 flux, θ is a vector of parameters, specifically V_{max} (the maximum reaction rate given saturation of either C, in forward MM, or microbial biomass (MB), in reverse MM), k_h (the half-saturation constant), and $1-\epsilon$ (accounting for carbon-use efficiency, ε).

Our specific model for F is:

$$F = (1-\varepsilon) \times C \times \frac{V_{max} MB}{k_h + MB} \tag{3}$$

Following the terminology of Chesson (1998,2012), the above is our "patch" model and our goal is to understand how spatial variances and covariances impact the integrated flux, which represents the spatial expectation or E[F] (hereafter denoted $\overline{F}$), which represents $\overline{-C \times (1-\varepsilon) \times \frac{V_{max} MB}{k_h + MB}}$.

The incorrect approach to solving for E[F] is to simply plug-in the mean-field solution:

$$\overline{-C} \times \overline{(1-\varepsilon)} \times \frac{\overline{V_{max}}\ \overline{MB}}{\overline{k_h} + \overline{MB}} \tag{4}$$

Analytically, an exact solution would require specification of a joint distribution for C and MB $\pi(MB, C, \theta)$, and solution of the convolution integral:

$$\int -f(MB, C, \theta)\pi(MB, C, \theta) dMB dC d\theta \tag{5}$$

However, following Chesson (2012) and Chakrawal et al. (2019) we are free to approximate the solution for arbitrary distributions using a Taylor Series approximation expanded to the 2nd moment. Specifically, we take the expectation over a multivariable Taylor Series expansion, centered around the mean-field values of all parameters θ (for simplicitiy, the variables MB and C are included in the parameter vector θ):

$$\overline{F} \approx \mathrm{E}\left[f(\overline{\theta}) + \frac{1}{2}\theta_{\theta-\overline{\theta}}^{T} H_{\overline{\theta}}[f(\theta)]\theta_{\theta-\overline{\theta}}\right] \tag{6}$$

where $H[f(\theta)]$ represents the Hessian matrix of the function that determines the CO_2 efflux F (in this case Michaelis-Menten), $\theta_{\theta-\overline{\theta}}$ represents the deviation from the mean at each instance and for each of the parameters. It can easily be seen that $\theta_{\theta-\overline{\theta}}^{T} \times \theta_{\theta-\overline{\theta}}$ is the variance-covariance matrix, and that the first moment of the Taylor expansion cancels because the first derivative of $\theta_{\theta-\overline{\theta}}$ is zero.

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

2.1 Non-dimensionalization

Expanding equation 6 out, we have 5 terms involving the variances of C, MB, $1-\varepsilon$, V_{max}, and k_h, and 10 terms involving covariances among the parameters. We can redistribute the expectation operator over this approximation to see that we are dealing with the contributions from the variance-covariance terms, weighted by the second partial derivatives evaluated at the mean for each parameter. However, the resulting expression **does not readily yield insight into the impact of scale transition upon the dynamics**, since second partial derivatives and cross partial derivatives do not have easy intuition. Morever, variances and covariances depend arbitrarily upon the scale of units and measurements involved, hindering both intuition and cross-site comparisons. Therefore, we non-dimensionalize equation 6 for $\mathrm{E}[F]$ as follows:

1) We define a dimensionless quanity λ as $\frac{\overline{MB}}{\overline{k_h}}$. λ thus represents a multiplicative factor expressing the ratio of the mean microbial biomass over it's half-saturation value, indicating the microbial saturation for the decomposition.
2) We divide all of the terms in 6 by their mean-field value, and represent the whole equation as a product:

$$\overline{F} \approx f(\overline{\theta}) + f(\overline{\theta})((\theta-\overline{\theta})^T \frac{\frac{\partial^2 f}{\partial\theta^2}_{\theta=\overline{\theta}}}{f(\overline{\theta})}(\theta-\overline{\theta})) = f(\overline{\theta})(1+\left(\theta-\overline{\theta})^T \frac{\frac{\partial^2 f}{\partial\theta^2}_{\theta=\overline{\theta}}}{f(\overline{\theta})}(\theta-\overline{\theta})\right)$$

(7)

3) We calculate the resulting expression for $\overline{F}$
4) We notice that $\frac{Var(\theta)}{\overline{\theta}^2}$ can be re-expressed as $(\frac{SD(\theta)}{\overline{\theta}})^2$ which in turn is the square of the dimensionless coefficient of variation $(CV(\theta))^2$. This enables us to reformulate the variance terms in (7).
5) Similarly, since the covariance terms can be rewritten as $COV(X,Y) = \rho_{X,Y}SD(X)SD(Y)$, we have the following equality:

$$\frac{COV(X,Y)}{\overline{XY}} = \rho_{X,Y}CV(X)CV(Y) \tag{8}$$

Applying steps 1-5 to all the terms in the equation, we end up with a fully dimensionless equation:

$$\overline{F} \approx f(\overline{\theta})(1+\frac{\lambda}{(1+\lambda)^2}[\rho_{k_h,MB}CV(k_h)CV(MB) - CV(MB)^2] + \frac{1}{(1+\lambda)^2}[CV(k_h)^2 - \rho_{k_h,MB}CV(h)CV(MB)] + \frac{1}{(1+\lambda)}[\rho_{C,MB}CV(C)CV(MB) + \rho_{V_m,MB}CV(V_m)CV(MB) + \rho_{\epsilon,MB}CV(\epsilon)CV(MB) + \rho_{C,k_h}CV(C)CV(k_h) - \rho_{V_m,k_h}CV(V_m)CV(k_h) - \rho_{k_h,\epsilon}CV(k_h)CV(\epsilon)] + \rho_{V_m,C}CV(V_m)CV(C) + \rho_{C,\epsilon}CV(C)CV(\epsilon) + \rho_{\epsilon,V_m}CV(\epsilon)CV(V_m) \tag{9}$$

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

Note that by symmetry, we have also solved for the case of the forward Michaelis-Menten kinetics. This can be expressed simply by interchanging C and MB, and by correspondingly altering λ to represent the ratio of substrate availability over half-saturation.

3. Discussion

Having fully non-dimensionalized equation 6, we are in a much better position to gain analytical insight into the scale transition. To begin, we note the pivotal role played by the quantity λ throughout this equation. λ scales the contributions of the parameter variation and correlation terms to the deviation from mean field behavior according to the ratios $\frac{\lambda}{(1+\lambda)^2}$, $\frac{1}{(1+\lambda)^2}$, and $\frac{1}{1+\lambda}$. In particular, all of the parameter variance terms (which have become $CV(\theta)^2$ upon non-dimensionalization), are scaled by one of these three λ ratios, alongside 7 out of 10 of the covariance terms. Overall, low λ (here $\lambda <\approx 1$) keeps all of the spatial correction terms in play, while increasing λ tends to simplify matters. As noted by Sihi et al. (2018), as $MB \to \infty$ (equivalent to MB >> k_h or $\lambda \to \infty$), reverse Michaelis-Menten kinetics converge to first order, leaving:

$$\overline{F} = -C \times (1-\epsilon) \times V_{max} \tag{10}$$

Accordingly, in our setup, the multiplicative factor for the scale transition correction approaches a simplified expression, as $\lambda \to \infty$:

$$\overline{F} \to f(\overline{\theta})\left(1 + \rho_{V_m,C}CV(V_m)CV(C) + \rho_{C,\epsilon}CV(C)CV(\epsilon) + \rho_{\epsilon,V_m}CV(\epsilon)CV(V_m)\right) \tag{11}$$

This is actually quite remarkable. Despite invoking the situation where microbial biomass (and its enzyme supply) is effectively infinite - thus linearizing the underlying patch models - we cannot eliminate the possibility of a potentially substantial deviation from mean-field when scaling decomposition kinetics. We note that in this resulting expression, we have reduced the situation to a set of three critical correlations involving two microbial physiological parameters (ϵ, and V_m), and substrate availability (C). Regardless of their respective variabilities (CV terms), if these correlations are close to zero, then the whole expression converges to mean field.

Returning to the situation where λ is not large, if we ignore the correlation terms (temporarily setting to zero), we see that there are direct contributions to the scale transition from the variability in MB and k_h that may, to some extent, balance each other:

$$\overline{F} = f(\overline{\theta})(1 + \frac{1}{((1+\lambda)^2)}[CV(k_h)^2] - \frac{\lambda}{(1+\lambda)^2}[CV(MB)^2]) \tag{12}$$

Focusing on the offsetting correction terms, we can re-write as:

$$\frac{\lambda}{(1+\lambda)^2}[\frac{CV(k_h)^2}{\lambda} - CV(MB)^2] \tag{13}$$

and for the case of $\lambda = 1$, this becomes:

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

$$\frac{1}{4}[CV(k_h)^2 - CV(MB)^2] \tag{14}$$

Thus, variability in the factors of soil protection that impact upon k_h in practice, can offset the impact of variability in microbial biomass itself.

More generally, starting with our dimensionless equation 9 puts modelers and empiricists in a better position to assess the quantitative significance of the scale transition correction across systems compared to expressions with opaque second partial derivatives and cross derivatives, and arbitrarily scaled variance terms. By re-expressing $\overline{F}$ in terms of dimensionless coefficients of variation, correlation coefficients and λ, we can plug-in realistic values for variability in any relevant parameter and assess the % effect on $\overline{F}$ in terms of deviation from mean field behavior. We argue that this formulation possesses significant advantages not only in understanding how to scale flux estimates ($\overline{F}$) *within* a site, but going forward will help facilitate intercomparison *among* sites in terms of their scale-free variability.

3.1 Spatial Colocation of Microbes and Substrate

To illustrate these advantages in interpretability, we first take the special case of a model where we treat all parameters as constant (and known) except substrate and microbial biomass. This corresponds to setting the other CV and ρ terms to 0. In this case, we are isolating the impact of the spatial colocation of substrate and decomposers. Our equation becomes:

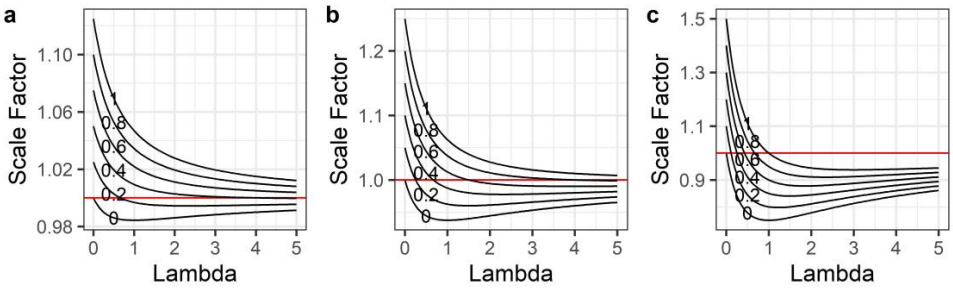
$$\overline{F} \approx f(\overline{\theta})(1 - \frac{\lambda}{(1+\lambda)^2}CV(MB)^2 + \frac{1}{(1+\lambda)}(\rho_{C,MB}CV(C)CV(MB))) \tag{15}$$

In the case of this formulation, there is a very clear dual convergence asλ increases:

1. deviation from mean-field behavior declines, and
2. first order kinetics are approached

Indeed, our equation 15 reveals the exact speed of this convergence in terms of dimensionless λ and a balance of $CV(MB)$, $CV(C)$ and their correlation.

We illustrate the scale transition solutions to equation 15 as a function of lambda for various choices of CV(C), CV(MB) and ρ:



https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

Figure 1: Scale correction factor for models given spatial colocation between microbes and substrate across a gradient of λ values, and for a variety of correlation ρ values (0-1), with CV(SOC) held constant at 0.5, a) CV(MB) = 0.25, b) CV(MB) = 0.5, and c) CV(MB) = 1.

We see in Fig. 1 that in the case of pure spatial colocation, with no variation in the kinetic parameters, the scale transition correction factor varies from a maximum of 1.5 to a minimum around 0.75, and in all cases indeed converges to 1 as lambda increases. The variability assumed for C and MB impacts only the scale of the correction factor, not the qualitative behavior as λ and ρ vary. One virtue of having a simplified, generic dimensionless equation of this sort is that it enables us to think in a unit-free/scale-free manner about the *plausible* range of the scale transition correction given transparent assumptions about variability and correlations.

Another virtus is that it is mathematically tractable to see how the variance and covariance terms can balance each other, and to solve for where they are equal. If we introduce a new term λ_2 representing the relationship between $CV(MB)$ and $CV(C)$ as follows $CV(MB) = \lambda_2 CV(C)$, we can re-express the the deviation of the mean-field correction from 1 as:

$$CV(MB)^2[\frac{1}{1+\lambda}\left(\rho\lambda_2 - \frac{\lambda}{(1+\lambda)}\right)] \tag{16}$$

Thus, whether the correction is positive or negative depends crucially on the product of the colocation correlation coefficient ρ and the extent of variability in substrate relative to variability in microbes.

If we fix λ_2 to unity, as done in our Fig.1, our mean-field deviation simplifies to:

$$CV(MB)^2[\frac{1}{2}(\rho\lambda_2 - \frac{1}{2})] \tag{17}$$

In general, it is clear that the scale transition correction is larger to the extent that microbial variability exceeds substrate variability under reverse michaelis-menten kinetics (the opposite relation holds for forward Michaelis-Menten by symmetry). Thus, variability in microbial biomass is not only important in itself in driving Jensen's Inequality, but also with respect to variability in substrate supply. **Our analysis thus highlights another route of convergence back to the mean field beyond the simple increase of λ: variability in substrate increasing to match variability in microbes in the presence of positive spatial colocation factor**. We also note that the magnitude of the mean field correction scales as the *square* of the coefficient of variation of microbial biomass. Quadratic scaling means that at low to moderate levels of variability, the deviation from mean field behavior is likely to be fairly minimal, but at moderately high to high levels of variability, severe deviations can be expected.

3.2 Environmental Heterogeneity

So far we have analyzed in depth the role of variability in microbes and their substrate, but not in the ecological drivers underlying maximal reaction rates (i.e. V_{max}) or half-saturation (i.e. k_h). We start with the observation that both linear first order and non-linear microbial models will show characteristic scale transitions given heterogeneity in temperature and soil moisture. Consider the asymptotic convergence of the reverse MM to first order

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

$$\frac{dC}{dt} = -V_{max}\overline{C} \tag{18}$$

This is mathematically equivalent to the more standard way of writing these models down as

$$\frac{dC}{dt} = -kC \tag{19}$$

To make matters clear, we re-express the rate limiting maximal reaction velocity V_{max} first as a function of temperature (assuming all else constant):

$$V_{max} = e^{aT} \tag{20}$$

In this case, our integrated flux equation will be:

$$\overline{\frac{dC}{dt}} = \overline{-e^{aT} \times (1-\epsilon) \times C} \tag{21}$$

Allowing for variability in T, this integrated equation will show characteristic scale transitions given the convex (exponential) relationship with T.

Using the Taylor expansion again to second order we have:

$$\overline{V_{max}} \approx e^{a\overline{T}}(1 + \frac{1}{2}a^2 Var(T)) \tag{22}$$

The critical scale correction term here is again multiplicative, and we re-express it into a function of a dimensionless coefficient of variation parameter more suited to ready interpretation. First, the exponential dependence of respiration on temperature is canonically codified in terms of Q_{10} scaling. We substitute $a = \frac{log(Q_{10})}{10}$, and end up with:

$$1 + \frac{1}{2}a^2 Var(T) = 1 + \frac{1}{2}(\frac{log(Q_{10})}{10})^2 Var(T) = 1 + \frac{1}{200}(log(Q_{10}))^2 (SD(T))^2 = 1 + \frac{1}{200}(log(Q_{10}))^2 (\overline{T}CV(T))^2 \tag{23}$$

For a "typical" Q_{10} of 2.5, and a $\overline{T}$ of 25, we see the multiplicative scale transition correction in figure 2:

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

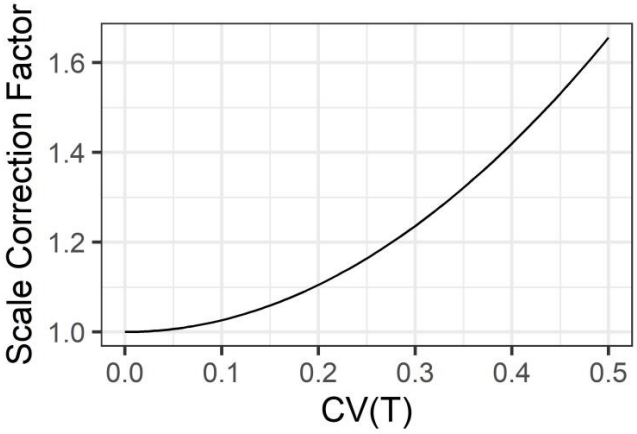


Figure 2: Scale correction factor for Q10 temperature response scaling given coefficient of variation CV(Temp) from 0 to 0.5

As is clear in Fig. 2, the scale transition for temperature is extremely convex. Integration of fluxes over ecosystems with significant heterogeneity in temperature invokes substantial deviation from a mean-field model. For instance, at a CV of 0.2, the scale correction factor is 1.10, but by a CV of 0.5 it is 1.66. Obviously, the significance of this depends on the scale and heterogeneity over which an accurate flux model is desired. For a smaller footprint eddy covariance tower (nuria) over a uniform habitat type, soil (and near surface) temperatures probably do not vary by much more than 20%. Regardless, our general mathematical analysis quantifies and clarifies exactly how the scale of variation influences the degree of the scale transition correction.

Notably, the only difference between the scale transition correction for first order and for reverse MM kinetics, is that in the latter, there would be additional correlation terms to consider, e.g. the correlation between temperature and V_{max}, temperature and k_h, as well as temperature and C and MB.

3.3 Lessons for Scientific Inference

We close our discussion by considering the implications of the scale transition for advancing the state of biogeochemical modeling. Critically, the representation of non-linear (microbial driven) kinetics is a crucial modeling choice with large implications for long-term SOC forecasts. Traditional first-order PBMs dodge explicit representation of these kinetics, but nonetheless have worked well in practice. This state of affairs persists because both non-linear and linear kinetics are capable of representing coarse-scaled biogeochemistry reasonably well, at least in certain respects. Since first order kinetics are known to be a crude approximation, the crucial question for practice is not whether they are “true”, but rather whether there is significant, systematic information loss inherent to their use. Fortunately, the scale transition offers a clear, clean path to discriminate between these alternative model formulations.

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

As noted throughout, the dimensionless term λ plays a critical role in linking the non-linear (michaelis-menten) kinetics to the first order kinetics. As λ increases, the non-linear kinetics converge to first order. Thus, in seeking to infer where the non-linear kinetic models provide substantial advantages, ensuring that λ is not too large (>>1) is the first priority. Previous work Sihi et al. (2016) has approached this question theoretically, from first principles. Here, we point out that demonstrating substantial deviation from mean-field model when fitting non-linear kinetics to data is both a necessary and sufficient condition for infering that λ is not too large. Thus we recommend that time series of flux data be fit to both a first order and a non-linear kinetic model, where crucial covariates including subtrate (SOC), microbial biomass, and possibly environmental parameters such as temperature, have been measured sufficiently well to quantify the relevant variances and covariances. Where predictive performance and forecasting are the primary goals, we recommend careful consideration of model parameterizations (i.e. based on leave-one-out cross validation), and model combination via "stacking" where it is difficult to infer a decisive "winner" (Yao et al. 2018).

In addition to the role of λ, our analysis also cleanly shows the contribution of other terms to the scale transition, and thus alternative metrics to assess. First and foremost, accounting for the spatial colocation of microbial biomass and substrate (according to equation 15 above) or the various correlation terms between microbial biomass and kinetic/environmental factors in equation 9. In addition to fitting fully parameterized flux models (as above), simpler statistical models could be fit examining the role of variations in microbial biomas, or colocation of microbial biomass and SOC, in explaining across-site variations in ecosystem respiratory fluxes (F). **A substantial role for either correlation of MB and C, or their variability, would constitute ipso facto evidence of the preferability of well-formulated non-linear kinetic models**. On the other hand, small roles for colocation, or evidence of large values of λ in practice would suggest minimal advantage to abandoning first order models in favor of more complex microbial models. A meta-analytical approach across sites will benefit greatly from our formulation in terms of dimensionless quantities like λ and the various coefficients of variation.

4. Conclusions

Here, we have illustrated how the spatial scale transition can be expressed in dimensionless form, yielding insight into the systematic operation of Jensen's Inequality in upscaling microbial decomposition kinetics. Our analysis has identified the central role of the dimensionless quantity λ - representing the ratio of mean-field microbial biomass over its half-saturation value - in governing the extent of the scale transition correction, expressed here in multiplicative form best facilitating comparison among systems. For somewhat simplified scenarios - such as restricting to spatial colocation of substrate and microbes - as $\lambda \to \infty$, the mean-field correction goes to 0 and the model converges to first order.

This dual sense of convergence also provides opportunity to empirically test for the presence of significant non-linear microbial dynamics in upscaled field data: to the extent that upscaled fluxes deviate from the flux estimated at mean-field conditions, we have *ipso facto* evidence for the importance of formulating our biogeochemical models with these non-linear terms. Conversely, where there is close agreement between mean-field and upscaled fluxes, there are arguably stronger reasons for retaining first-order process model formulations.

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

In closing, we would like to point out how this mathematical analysis illustrates the challenge of scaling quite nicely. In the context of non-linear models, for each parameter that is allowed to vary in space, there is not only a new variance parameter, but a number of new covariance terms are induced, growing as the factorial of the number of varying parameters $\binom{5}{2}!$ (Fig. 3). Thus, in the case of the 5 parameter function considered here, the full approximation has 5 mean field terms, 5 coefficients of variation, 10 correlation coefficients, and the dimensionless quantity λ.

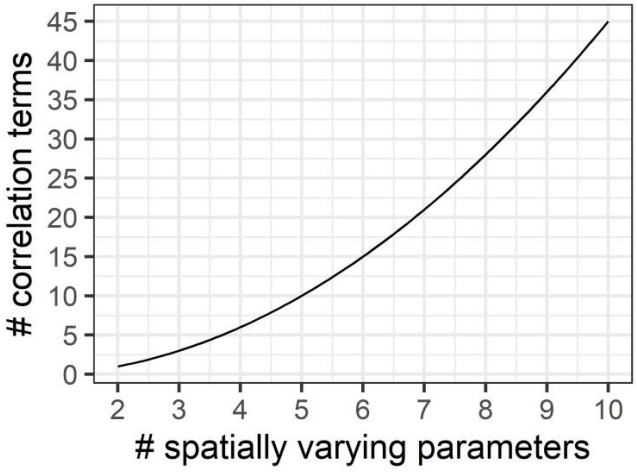


Figure 3: Model complexity grows exponentially with number of spatially varying parameters. We argue to keep models as simple as possible for both analytical and computational tractability.

Even with a maximally generic and simplified expression, fitting such non-linear time series models to field data still represents quite a challenge, especially while adequately accounting for and propagating uncertainty. Modelers and theoreticians should appreciate the complexity of the task at hand. Fortunately, our analysis has identified a potentially robust route to limiting model complexity: screen systematically for the importance of various correlations in explaining variations in fluxes. Accordingly, we recommend that research focus first upon spatial colocation of MB and C, which is readily measured, and then to thoughtfully and carefully expand models with additional terms as needed.

Author Contributions: CHW conceived the original concept, developed the mathematical analysis, and wrote the manuscript. SG developed the concepts, contributed to the mathematical analysis, and co-authored and edited the manuscript.

Competing Interests: None declared.

Acknowledgements: We stand on the shoulders of giants: Peter Chesson's research program on Scale Transition was enormously influential. We thank T. Trevor Caughlin for introducing us to Chesson's papers many years ago, and to everyone who has humored discussions of Jensen's Inequality ever since!

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

References

Blankinship, Joseph C., Asmeret Asefaw Berhe, Susan E. Crow, Jennifer L. Druhan, Katherine A. Heckman, Marco Keiluweit, Corey R. Lawrence, et al. 2018. "Improving Understanding of Soil Organic Matter Dynamics by Triangulating Theories, Measurements, and Models." *Biogeochemistry* 140 (1): 1–13. https://doi.org/10.1007/s10533-018-0478-2.

Blankinship, Joseph C., and Joshua P. Schimel. 2018. "Biotic Versus Abiotic Controls on Bioavailable Soil Organic Carbon." *Soil Systems* 2 (1): 10. https://doi.org/10.3390/soilsystems2010010.

Chakrawal, Arjun, Anke M Herrmann, Johannes Koestel, Jerker Jarsjö, Naoise Nunan, Thomas Kätterer, and Stefano Manzoni. 2019. "Dynamic Upscaling of Decomposition Kinetics for Carbon Cycling Models." *Geosci. Model Dev. Discuss. Doi* 10.

Chesson, Peter. 1998. "Spatial Scales in the Study of Reef Fishes: A Theoretical Perspective." *Australian Journal of Ecology* 23 (3): 209–15. https://doi.org/10.1111/j.1442-9993.1998.tb00722.x.

———. 2009. "Scale Transition Theory with Special Reference to Species Coexistence in a Variable Environment." *Journal of Biological Dynamics* 3 (2-3): 149–63. https://doi.org/10.1080/17513750802585491.

———. 2012. "Scale Transition Theory: Its Aims, Motivations and Predictions." *Ecological Complexity*, From spatially explicit population models to mean-field dynamics, 10 (June): 52–68. https://doi.org/10.1016/j.ecocom.2011.11.002.

Georgiou, Katerina, Rose Z. Abramoff, John Harte, William J. Riley, and Margaret S. Torn. 2017. "Microbial Community-Level Regulation Explains Soil Carbon Responses to Long-Term Litter Manipulations." *Nature Communications* 8 (1): 1223. https://doi.org/10.1038/s41467-017-01116-z.

Greenland, Sander. 2019. "Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution with S-Values." *The American Statistician* 73 (sup1): 106–14. https://doi.org/10.1080/00031305.2018.1529625.

Levin, Simon A. 1992. "The Problem of Pattern and Scale in Ecology: The Robert H. MacArthur Award Lecture." *Ecology* 73 (6): 1943–67. https://doi.org/10.2307/1941447.

Parton, W. J., D. S. Schimel, C. V. Cole, and D. S. Ojima. 1987. "Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands." *Soil Science Society of America Journal* 51 (5): 1173. https://doi.org/10.2136/sssaj1987.03615995005100050015x.

Ross, Sheldon. 2002. *A First Course in Probability*. Pearson Education India.

Schimel, Joshua P., and Michael N. Weintraub. 2003. "The Implications of Exoenzyme Activity on Microbial Carbon and Nitrogen Limitation in Soil: A Theoretical Model." *Soil Biology and Biochemistry* 35 (4): 549–63. https://doi.org/10.1016/S0038-0717(03)00015-4.

https://doi.org/10.5194/bg-2021-108
Preprint. Discussion started: 19 May 2021

Sihi, D., S. Gerber, P. W. Inglett, and K. S. Inglett. 2016. "Comparing Models of Microbial–Substrate Interactions and Their Response to Warming." *Biogeosciences* 13 (6): 1733–52. https://doi.org/10.5194/bg-13-1733-2016.

Todd-Brown, Katherine E. O., Bin Zheng, and Thomas W. Crowther. 2018. "Field-Warmed Soil Carbon Changes Imply High 21st-Century Modeling Uncertainty." *Biogeosciences* 15 (12): 3659–71. https://doi.org/10.3929/ethz-b-000272804.

Van Oijen, Marcel, David Cameron, Peter E. Levy, and Rory Preston. 2017. "Correcting Errors from Spatial Upscaling of Nonlinear Greenhouse Gas Flux Models." *Environmental Modelling & Software* 94 (August): 157–65. https://doi.org/10.1016/j.envsoft.2017.03.023.

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. 2018. Using stacking to average Bayesian predictive distributions (with discussion). *Bayesian Analysis*, *13*(3), 917-1007.

Wieder, William R., Steven D. Allison, Eric A. Davidson, Katerina Georgiou, Oleksandra Hararuk, Yujie He, Francesca Hopkins, et al. 2015. "Explicitly Representing Soil Microbial Processes in Earth System Models." *Global Biogeochemical Cycles* 29 (10): 2015GB005188. https://doi.org/10.1002/2015GB005188.

Wieder, William R., Gordon B. Bonan, and Steven D. Allison. 2013. "Global Soil Carbon Projections Are Improved by Modelling Microbial Processes." *Nature Climate Change* 3 (10): 909–12. https://doi.org/10.1038/nclimate1951.

Wieder, William R., Melannie D. Hartman, Benjamin N. Sulman, Ying-Ping Wang, Charles D. Koven, and Gordon B. Bonan. 2018. "Carbon Cycle Confidence and Uncertainty: Exploring Variation Among Soil Biogeochemical Models." *Global Change Biology* 24 (4): 1563–79. https://doi.org/10.1111/gcb.13979.