Universal Curves to Represent Heterogeneity in Soil Carbon Residence Times

Chris H Wilson, Stefan Gerber

10/19/2020

# Background

Soil organic carbon consists of a heterogeneous assortment of organic molecules phyisically protected from microbial decomposition through soil mineral interactions. This heterogeneity has traditionally been represented in process-based models (PBMs) using multiple discrete pools differentiated on the basis of residence time, e.g. the labile, slow and passive pools in CENTURY (Parton 1988), or by differing mechanisms of protection, e.g. the physical versus biochemical pools in MIMICS (Wieder 2014). Both approaches are zero order approximations of a necessarily far richer reality. For this reason, efforts to operationalize these various “pools” in the form of empirical procedures (e.g. size-density fractionation) have not generally been satisfying. Stable isotope probing has suggested that any such operationally defined “pool” is itself heterogeneous (citations: our work at BRU, others). Recently, Waring et al. (2020) propose an individual-based model (IBM) framework in which cohorts of carbon atoms are tracked probabilistically over time, as they transform from one biochemical state to another, and participate in various soil mineral interactions. One happy feature of this approach is that residence time is correctly represented as an *emergent property*, rather than identified as an inherent feature of a supposedly homogeneous pool. In fact, the output of the Waring et al. (2020) model is a *distribution over residence times*, for each major class of organic molecule, thus providing one promising path to embracing real-world SOC heterogeneity. However, regardless of **where** the heterogeneity comes from, or **how** it is represented, there is a clear need for a consistent mathematical framework that will allow us to deduce the consequences of this heterogeneity for long-term SOC dynamics, and prediction of the ecosystem services that depend on SOC, such as greenhouse gas mitigation. Below, I outline a generic, quantitative framework that maps a distribution over SOC residence times to a scaled-up dynamics via use of an integral transform. I derive two universal curves bounding the possible scaled-up dynamics, and provide a clear path to map from fitted models to ecosystem service.

# Scaling up over SOC heterogeneity: integral transform

Given a distribution () of residence times () for a given class of organic molecules within a soil, the contribution of a cohort of carbon atoms to offseting atmospheric carbon is simply the fraction remaining in the soil at time . Rather than rely on expensive computational experiments, this can be solved analytically, with tremendous payoff. We represent the scaled-up dynamics as an expectation over the heterogeneity:

(1)

which turns out to be essentially the Laplace transform of the random variable .

In short, the framework proposed here invokes two assumptions. First, our cohort of carbon atoms is sufficiently large that we can apply a “mass action” assumption. That is, we represent their possible fates by the residence time distribution convoluted with turnover via the exponential in equation (1). Basically, this is a simple first order decay assumption at the level of the individual, while *the behavior of the ensemble* requires the integral transform in (1). Secondly, we start with the observation that the heterogeneity in SOC residence times (Waring et al. 2020) is generally right-skewed and unimodal, and we capture these features via an inverse-gamma distribution with mild restrictions.

With these assumptions in place, the specific integral transform is:

(2)

Fortunately, this integral has a very simple analytical solution, dropping the cohort constant “C”:

(3)

We reparameterize this solution because, while simple, the parameters of the inverse-gamma distribution and do not necessarily have easy intuition. However, we can translate into a parameterization with a mean and a dispersion constant , and rewrite the integral transform in (3). To do this, we first note that the first moment (mean) of an inverse gamma is given by , while the second moment (variance) is . The first moment is only defined when >1, and the second moment only when >2. Thus, for our reparameterization, we will set the restriction that is strictly > 2, essentially asserting that the distribution of residence times belongs to a right-skewed class of distributions with finite mean and variance. We notate the first moment as , and retain this parameter. For second moment, note that:

(4)

So, we define a new parameter . For interpretation, turns out to be equivalent to the coefficient of variation (CV) squared, and thus we consider a variability or dispersion constant. Solving for the parameters and substituting, the integral transform in (3) is now:

(5)

To be extremely clear, the amount of carbon from a given cohort of carbon atoms remaining at time is given by (6):

(6)

Thus, given a time series of observations of turnover (e.g. from incubation experiments such as XXXX), we could estimate the relevant parameters governing the distribution of residence times and inversely by fitting to equation (6). Alternately, we can use forward simulation results from a suitable PBM that provides such a distribution (e.g. Waring et al. 2020), plug in the resulting parameters and use (6) to estimate scaled-up properties and services.

# Universal bounding curves

The function in (6) can be further probed for theoretical insight by non-dimensionalization and examination of limiting behavior. First, we consider the case where variablity becomes maximal, by taking the limit , resulting in:

(7)

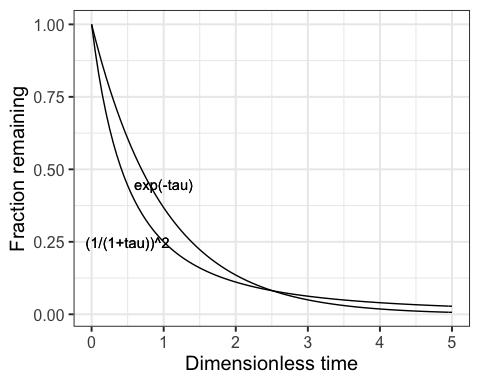
We non-dimensionalize 7 with the quantity , where is thus a dimensionless time representing multiples of the mean of the residence time distribution, and we have:

(8)

From the other direction, we let , and find that (6) is equivalent to the better known limit:

(9)

which of course is equivalent to a first-order representation of turnover with a fixed mean residence time, and we nondimensionalize with again.

In Figure 1 below, we show the universal bounding curves: 

*Fig 1: Fraction of carbon remaining in a given class given heterogeneity in turnover as a function of dimensionless time, for the case of maximal variance (1/(1+tau))^2, or of no variance, which is the limiting case of exp(-tau).*

As is evident in figure 1, while holding the mean of the residence time distribution constant, we can generate qualitatively distinct behavior by varying the amount of dispersion (). Greater dispersion leads to the behavior that the initial loss or turnover is much steeper, but in the long run, a larger amount remains in the system. We thus gain insight into the behavior that has frequently motivated fitting of two or three pool models to incubation or turnover data, where the interpretation is that there are distinct “labile” versus “recalcitrant” pools. **Instead, variability in a unimodal distribution of residence times can itself lead to this behavior**. Invocation of distinct labile and recalcitrant pools may thus be doubly an error. However, the possible turnover dynamics are very well-constrained by these two curves in dimensionless (hence universal) space.

# Discussion and Application to Ecosystem Service Quantification

Regardless of the source of heterogeneity in a pool or cohort of SOC, the equation in (6) represents the fraction of C remaining at some arbitrary time given only that a sufficiently large cohort of carbon atoms are involved, and that the distribution of residence times is unimodal and right-skewed. We have shown through analysis of dimensionless forms that the range of behaviors is bounded by the two curves in Figure 1. Moreover, variability in residence times *per se* can generate the kind of qualitative behavior associated normally with distinctions between labile/recalcitrant pools. The equation in (6) could readily be fit to turnover or gas efflux data from incubation or isotope pulse-chase experiments to estimate the parameters. An open question is: how long does data need to be collected to adeuqately constrain and , since those parameters in turn govern the long-term service?

If instead a process-based model, such as the PROMISE model in Waring et al. (2020), gives us a distribution over residence times, we can easily compute the service of *avoided greenhouse forcing*, by taking the integral of (6) over time, yielding:

(10)

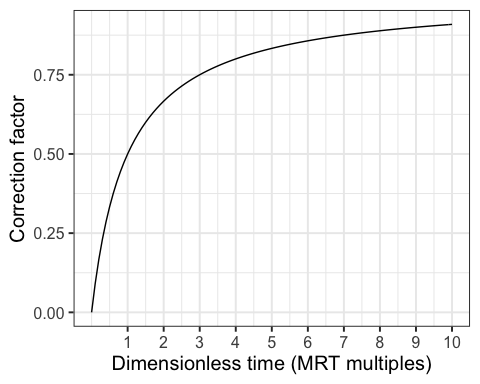
Remarkably, in the limit as , the ratio on the RHS goes to 0, and the whole estimation of the service simply becomes:

(11)

where is the mean of the residence time distribution. Equation (11) is isomorphic to the solution for equilibrium carbon stocks under first order decomposition, if we were to interpret as a continuous input, rather than a single cohort, thus demonstrating continuity between our framework and equilibrial analysis. Ironically, given an infinite time horizon, the impact of SOC heterogeneity washes out and essentially the service estimation only requires robust estimation of the first moment (mean). For finite time horizons, we need the correction term and the full equation in (10), and thus the variability/dispersion parameter enters again. To generate a universal bounding curve we once again take the limit as , and non-dimensionalize the ratio, resulting in:

(12)

where is dimensionless time (multiples of ), resulting in Figure 2 below.



*Figure 2: Bounding curve for the maximal heterogeneity correction factor in estimating how the integral of avoided GHG forcing, the standard method of quantifying Global Warming Power, varies with the time horizon considered as a multiple of the mean of the residence time distribution.*

The key insight from our analysis is this: increasing heterogeneity in residence times, while holding the mean of the distribution constant, leads to a higher long-term storage, but yet a lower service evaluated in terms of the GWP. The bounding curve in Figure 2 shows the correction to the service provision for the case of maximal variation. However, in the limit of an infinite time horizon, the avoided GWP turns out to just be the initial cohort size multiplied by the mean of the residence time distribution.

Given its evident importance to service quantification, how are we to interpret the meaning of the parameter ? As usual in applications of probability theory, we can consider to reflect underlying biophysical heterogeneity. However, we could also simply impart an epistemic interpretation to , as representing our **uncertainty** in residence times. In practice, with Bayesian inference we would work with a posterior distribution over , capturing both epistemic and biophysical cases. In both cases, we should think about , and the correction factor it induces (e.g. in Figure 2), as a sort of a “risk discount” in the provision of GWP service. Under high variability/uncertainty, the service estimate deviates - as it should - from the mean field estimate given by , especially in the short term.

# Conclusions

This reduces the computational and logistical burden of experimental and modeling work to only that sufficient to characterize the distribution over residence times, or perhaps a sum over a set of distribution of residence times (one for each class of molecules). Equation (6) is readily fit to empirical data, and thus to estimate the relevant parameters we have defined as and . From there, our mathematical framework presented here has analytically solved the scaling problem, and demonstrates clearly how heterogeneity in turnover impacts both the expected long-term dynamics and the quantification of ecosystem service provision. Happily, this behavior has particularly elegant mathematical form, and its limiting cases hve been presented in figures 1 and 2. Finally, we have derived an analytical expression (equation 10), that provides a “risk discount” for the impact of uncertainty in the provision of climate change mitigation by soil carbon sequestration, when we use avoided GWP as our basis.