Problems associated to kinetic fitting of incubation data

L. Weihermüller*[1], A. Neuser[2], M. Herbst[1], and H. Vereecken[1]

[1] Agrosphere Institute IBG 3, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany,

[2] Bergische Universität Wuppertal, Institute of General Geography, Human-Environment Research, D-42119 Wuppertal

*Corresponding author: Lutz Weihermüller, Forschungszentrum Jülich IBG-3 Agrosphere Institute D-52425 Jülich, Voice: ++49 (0)2461 61 8669 email: l.weihermueller@fz-juelich.de

Abstract

Incubation is a classical laboratory experiment to gain information about the carbon turnover from soils under different treatments (e.g., temperatures, water content, nutrient supply), amendments (e.g. biochar, straw, compost), or from different locations (e.g. topsoil vs. subsoil). Classically, the measured data are represented as cumulated CO_2 flux or as % total organic carbon (TOC) mineralized and from those data kinetic parameters are often derived using models of different complexity. Based on the conceptual idea that more than one C-pool exists, especially in soil mixtures, the simple single-pool model is more and more substituted by double-pool models to describe these data. Hereby, the model will be iteratively fitted to the data to obtain the pool sizes and rate constants of the pools (slow and fast). In the work presented, we show that the fitting of double-pool models will lead to different results in the pool sizes and kinetic parameters, depending on the fitting approach used. Secondly, general problems of over-fitting and the ill-posed problem are discussed, whereby it will be shown, that especially the estimation of the rate constant for the slow pool is highly uncertain. Based on these general findings and problems in the fitting procedure, fitting results reported in literature were analyzed. The meta-analysis indicates that only a small number of reported fits are apparently not well fitted, whereas a non-negligible percentage of reported fits were over-fitted or wrong parameters were reported. Finally, the paper provides guidelines for kinetic fitting and discusses possible fitting alternatives.

Introduction

The measurement of the production of CO_2 from soils in incubation studies has been used for many years to gain information about the influence of various soil types, changing temperatures and water contents, as well as the addition of amendments on the soil respiration (Reiners, 1968; Tayler & Parkinson, 1988; Mukherjee et al., 2015, amongst many others). Amongst the first who used incubation data in a qualitative mathematical description was Olsen (1963) who described the carbon turnover by the use of a single-pool model. In the following years mathematical modeling of the biodegradation kinetics became more and more important to obtain data that could be related to other studies and used for generalizing important findings. Additionally, such experiments were performed to derive the rate constants or half-lives of different carbon pools, especially of the labile carbon pool, as inputs into complex predictive carbon models such as RothC (Coleman and Jenkinson, 1996, 1999), Century (Parton et al., 1994), Candy (Franko et al., 1997), Daisy (Hansen et al., 1990) or Yasso (Liski et al., 2003, 2005). As listed, several models with different complexities exist to elucidate the carbon mineralization dynamics. Among them, multi-compartment models were designed to simulate the turnover of soil organic matter ranging from several days to 10,000 years, whereby the age of the recalcitrant carbon pool will be classically estimated by radio carbon dating (Trumbore, 1993). For incubation experiments the observation time, and therefore, the information content of the data, is far too limited to gain information about the recalcitrant pool with turnover times larger than several hundred years. As a consequence, mainly information about the labile and refractory pool will be derived from incubation data (e.g., Ajwa & Tabatabai, 1994; Liang et al., 2008; Qayyum et al., 2011; Saviozzi et al., 1997; Zimmerman et al., 2011). While in the early years kinetic modeling (or fitting) of incubation data was restricted to the single- or one-pool model due to the possibility of solving the problem by log-transforming the observed data and using a linear regression for the estimation of the rate constant, more recent publications chose multi-pool models (two-, three-, and even four-pool models), which are fitted iteratively using appropriate computer software. For the single-pool model the total organic carbon is assumed to decompose with a certain turnover rate, whereby for the double-pool model the total organic carbon is separated into two different compartments, i.e. into a labile pool with a short turnover time and into a non-labile or refractory pool with an accordingly longer turnover time (Qayyum et al., 2011). In some cases even pool models with more than two pools are reported and used to describe incubation experiments (e.g., Paul et al., 1999, Haddix et al., 2011; Li et al., 2013). In general, there are different fitting approaches used in literature to estimate the kinetic parameters from incubation studies, which will yield different kinetic parameters even for the same experimental data. This means in consequence, that the kinetic parameters fitted might be inconsistent. If these parameters will be used for comparison between studies or as inputs for predictive carbon modeling, different conclusions will be drawn or different model predictions will be generated. Unfortunately, literature data often does not allow to clearly identify which of the methods has been used and in some cases only refitting of the reported data will give insight into the approach applied.

A second drawback of kinetic parameter fitting is associated to the choice of the kinetic model applied (single-pool versus double-pool model). The motivation for the more frequent use of models with more than one pool can be found in the conceptual ideas developed over the last decades. The bulk soil organic carbon is not composed of one single substance, which turns over with a single specific rate, rather than a mix of different organic constituents exhibiting different qualities, and therefore, associated to different turnover rates. To account for this, the single-pool model was seen as being not state-of-the-art to describe the incubation data anymore, irrespectively of the information content included in the experimental data. In other words, it has to be questioned if the application of higher order pool models is suitable to receive a unique parameter set or if the fitting is ill-posed, which is defined as the drawback that the identified parameters are instable and/or non-unique. Non-uniqueness results from a non-convex objective function, which exhibits either multiple local minima or a global maximum that occurs for a broad range of parameter values. In that case, different combinations of parameters will provide the same model response (Šimůnek & Hopmans, 2002). This problem was already introduced as *equifinality* by Beven in the mid 1990 (Beven, 1993 and 1996) for hydrological modeling. Non-uniqueness means in consequence that different parameter combinations in the kinetic model will yield the same model response, and therefore, the estimated parameters, especially the rate constant, half-life or mean residence time, cannot be used for comparison between studies neither for initialization of complex predictive carbon models. Only little attempt was made to explore the applicability of higher order carbon pool models. For example, Scharnagl et al. (2010) analyzed the feasibility to estimate four pools of the RothC model (Coleman & Jenkinson, 1999) from incubation experiments using a Bayesian approach. Their results showed that the different pools could be adequately determined using a total incubation time of about 900 days. Li et al. (2013) used a Markov chain, Monte Carlo (MCMC) technique to estimate the kinetic parameters and to evaluate the identifiability and uncertainty of the estimated parameters. Even if they concluded that the kinetic parameters could be well estimated using the MCMC approach, a fairly large spread in the posterior distribution of the parameters is detectable (see Fig. 2 of their publication). This large spread in the fitted parameters is also reflected by the associated uncertainties in the reported C fractions for the three pools fitted with 4.7 ± 2.6 (mean ± SD), 22.4 ± 16.1 and 72.9 ± 17.6 %, respectively, and becomes even more evident in the calculated mean residence times of 0.19 ± 0.17, 2.71 ± 2.34 and 80.15 ± 61.14 years for pool 1, pool 2, and pool 3, respectively.

Based on the many drawbacks and problems associated with the estimation of kinetic parameters from incubation experiments this paper will i) give an overview of the theory of kinetic modeling ii) the problem of different data expression and approaches used for kinetic fitting, iii) provide information about drawbacks, especially over-fitting and the ill-posed problem, and will iv) analyze existing data from publications on the drawbacks introduced.

Background and Theory

Kinetic carbon model

In general, there are two different possibilities to express the experimental data from incubation experiments. The first is the expression as cumulative flux, where the single CO_2 measurements will be cumulated over time (see Fig. 1a). This approach will be denoted as *flux* in the following. The second possibility to express the experimental findings is to relate the measured CO_2 fluxes to the total organic carbon (TOC) content in the microcosm and to calculate the percentage of total carbon mineralized. This approach will be denoted as *% TOC mineralized* in the following. Again the % carbon mineralized will be cumulated over time as shown in Fig. 1b.

The general kinetic model to describe the data, irrespectively of the number of pools used and the way data are expressed (*flux* or *% TOC mineralized*), is given by:

$$C_t = (C_1 \times e^{-k_1 t}) + \ldots + (C_n \times e^{-k_n t}) \qquad [1]$$

where C_t is the mineralized total CO_2 *flux* (or *% TOC mineralized*) whereby C is also denoted as M in many studies. C_1 is the total size of the fast C-pool and C_n is the size of the slowest C-pool. k_1 to k_n are the corresponding rate constants [time units such as days^{-1} or years^{-1}] for the respective pools C_1 to C_n. In case of the data expression as *flux* C_t and C_n can be either expressed in CO_2 mass units [e.g., mg CO_2 g^{-1} soil] or as a percentage of the entire measured flux [%]. In some cases C_t and C_n will be expressed in C-equivalent calculated from evolved CO_2. Assuming a single-pool model reduces Eq. [1] to the first term on the right hand site and for a double-pool model n will be 2.

Different Constraints for the fitting

Additional to the two different possibilities to express the experimental data (*flux* or *% TOC mineralized*) different constraints can be used for the fitting of the kinetic parameters of the kinetic model (Eq. [1]).

I. *Unconstrained fitting*, where the fitting of the kinetic parameters from evolved CO_2 (or C equivalents) will be performed without constraining the total cumulative flux at t_{end}

II. *Constrained fitting*, where the fitting of the kinetic parameters from evolved CO_2 (or C equivalents) will be constrained to match the total cumulative flux (or *% TOC mineralized*) at t_{end}

Here it has to be noted, that for the second possibility (*constrained fitting*) pool sizes can be either expressed on mass basis (e.g., mg CO_2 g^{-1} soil) or in % from total flux, whereas for the first possibility (*unconstrained fitting*) only pool sizes expressed on mass basis (e.g., mg CO_2 g^{-1} soil) can be used.

Putting all possibilities together, four different fitting approaches are feasible and used in literature:

1. *Constrained fitting* on *flux* data using pool sizes expressed on mass basis (e.g., mg CO_2 g^{-1} soil), where C_1 to C_n sum up to total cumulative flux at t_{end}.
2. *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux, where C_1 to C_n sum up to 100% total cumulative flux at t_{end}.
3. *Unconstrained fitting* on *flux* data using pool sizes expressed on mass basis (e.g., mg CO_2 g^{-1} soil), where C_1 to C_n will sum up to any value.
4. *Constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized, where C_1 to C_n sum up to the total % TOC mineralized at t_{end}.

In approach 1, the kinetic parameters k_n and pool sizes C_n will be fitted for each C-pool n, whereby the sum of all pools n match the total flux at the end of the experiment, t_{end}. This can be also done by expressing the pool sizes C_1 to C_n as a ratio (%) of total flux observed leading to approach 2. In consequence the sum of the individual C-pools in both approaches (1 and 2) is always smaller than the total C-stock of the microcosm and reflects the mineralized carbon contributing to the total flux. In other words, the fitted kinetics (not only the pool sizes C_n but also the rate constant k_n) describe only the part of the soil carbon turned over (mineralized) and not the kinetics of the entire C-stock. Because only the mineralized C-pools will be analyzed, the fitted rate constants (k_n) are generally higher and the mean residence times

(MRTs) and half-lives (HL) are much shorter than those determined using approach 4. On the other hand, this approach has the advantage of providing kinetic parameters, which can be used for comparison of different soils / amendments / treatments in the same study. In approach 3, the kinetic parameters k_n and pool sizes C_n will be fitted for each C-pool *n*, whereby the sum of each estimated pool size will not match necessarily the total C-stock in the microcosm. The impact of the unconstrained fitting will be discussed in the *Results* section. In approach 4, the raw CO_2 flux data will be related to the total C-stock in the microcosms resulting in *% TOC mineralized* data. In contrast to the other approaches (1 to 3), the fitted kinetic parameters reflect the turnover of the entire C-stock in the microcosm and therefore, the rate constants fitted (especially for the fast or labile C-pool) can be used to inform complex predictive carbon models such as RothC, Century, Candy, or Yasso as well as to compare treatments within the study. The impact of the different fitting approaches on the kinetic parameters obtained and the implication for their use will be further discussed in the *Results* section.

Fitting of the kinetic model

Single-pool models were used since the 1960s to describe incubation data, whereby the fitting in the 1960s and 1970s was performed using an analytical solution, which was necessary because access to computers and fitting software was limited. Therefore, the raw data were log-transformed as shown by e.g. Stanford & Smith (1972), resulting in a linear regression between observed *flux* (or *% TOC mineralized*) and time. As mentioned, the resulting regression can be solved analytically for the rate constant. In this case, the rate constant is the regression slope of the log-transformed *flux (% TOC mineralized*) versus time. As an example, a single-pool model was used to generate cumulative *fluxes (% TOC mineralized*) over time with a rate constant k of 0.025 days^{-1} (see Fig. 2a). In a next step, the *fluxes (% TOC mineralized*) were log-transformed and plotted in Fig. 1b. As can be seen, all points arrange on a straight line, on which the regression and slope can be calculated. The back-calculated k value from the slope matches exactly the k value used as input. In the second example (Fig. 1c to d), a double-pool model was used to generate the *fluxes (% TOC mineralized*) with k_1 of 0.050 and k_2 of 0.002 days^{-1} and a pool ratio of 30 % for C_1 and 70 % for C_2. Again, these fluxes were log-transformed and a clear non-linearity is detectable, which reflects the underlying super-imposement of the two exponential functions. For better visualization, we also plotted two linear regressions into Fig. 1d but did not provide the regression equations due to two reasons. First, even if the kinetic parameters (k_1 and k_2) can be theoretically estimated by using two linear regression functions, the fitting is a crucial step. This is mainly based on the decision where to split the data set for the individual regressions. Based on the first problem, the obtained kinetic parameters often diverge from those parameters used to generate the data. To overcome the problem of the analytical solution based fitting of the double-pool model computers and appropriate fitting software was introduced from the mid-1980s.

Materials and Methods

Kinetic fitting, Statistical Analysis, and goodness of fit

For all fittings in this study, we followed the same procedure as described below.

The parameters providing the best prediction of the measured incubation data can be determined by minimizing the sum of squared residuals:

$$SSR = \sum_{i=1}^{m} (x_{obs,i} - x_{sim,i})^2 \qquad [2]$$

where x_{obs} and x_{sim} are the observed and simulated cumulative CO_2 (or C-equivalent) *fluxes* [e.g., mg CO_2 g^{-1} soil] or *% TOC mineralized* at time step *i* and *m* is the total number of observations. For the minimization of the objective function [Eq. 2] the rate constants k_n and the pool sizes C_n of Eq. [1] were systematically varied using the global optimization routine shuffled-complex-evolution University of Arizona (SCE-UA) as described by Duan et al. (1992 and 1994) until convergence was reached. The algorithm was considered to converge when the objective function changed <0.01% in 10 consecutive loops of the algorithm (1000–2000 model runs). This optimization routine has been already successfully applied in a wide range of applications in hydrology (Mertens et al., 2005; Mboh et al., 2011) but also for the estimation of parameters in non-linear C models (Weihermüller et al., 2009 & 2013; Bauer et al., 2012).

As a metric for the goodness of the fit the χ^2-test was used as recommended by FOCUS (2006), which is based on the degree of freedom *df*, computed as the number of data points *n* minus the number of model parameters *p*:

$$\chi^2 = \sum_{i=1}^{n} \frac{\left(x_{sim,i} - x_{obs,i}\right)^2}{\left(err/100 \cdot \bar{x}_{obs}\right)^2} \qquad [3]$$

where $\bar{x}_{obs}$ is the mean of all observed values and *err* is the measurement error percentage. In case χ^2 is larger than the tabulated $\chi^2_{df,\alpha}$ the model is not appropriate to represent the data at the chosen level of significance α. If different models will be tested (e.g., single versus double–pool model) the model with the lowest χ^2 should be selected. Here it has to be noted that many other statistic matrices can be used to identify the goodness of the fit, which are listed in Appendix 1.

Finally, the half-life (HL) and the mean residence time (MRT) can be calculated from the rate constants k_n by

$$HL = \frac{\ln(2)}{k_n} \qquad [4]$$

$$MRT = \frac{1}{k_n} \qquad [5]$$

Additionally to the statistical matrices a visual judgment of the goodness of the fit was (and should be) performed. Therefore, the measured data (*flux* or *% TOC mineralized*) and the calculated model curve was plotted. A second plot indicating the distribution of residuals (estimated minus measured data) over the time course was used. This residual plot is useful for revealing heterodescacity (systematic over- or under-predictions) and provides information about the bias. For an exact model fit, all residuals are zero. Systematic deviations occur if negative and positive residuals are not randomly scattered around the zero line. An example of ‘good’ and ‘bad’ fit is provided in Fig. 3 for some synthetically generated data. As can be seen, the good fit of the single-pool model (Fig. 3a) led to a random distribution of the residuals over time. On the other hand, Fig. 3b is an example where the single-pool model provides a poor fit to the data. This is obvious from visual inspection of both plots (model fit and residuals). As can be seen the fluxes near the beginning of the incubation are generally under-estimated by the model, resulting in relatively large positive deviations (ten consecutive positive residuals). Additionally, the last points are generally underestimated with relatively large negative residuals. In the case of such systematic errors the use of a single-pool model can be questioned and a double-pool model should be used instead. The help of the visual inspection is not only restricted to the application of the single-pool model but was (and should be) performed also for the fit of a double-pool model. Finally, the contributions of the single fluxes from the individual pool are also plotted in combination with the total flux to give impression of the pool contribution to the overall *flux* (*% TOC mineralized*).

Calculation of parameter spaces

Parameter spaces provide visual information about the identifiability of fitting parameters. The calculation of the parameter spaces for a double-pool model (Eq. [1] with n = 2) was explored with a grid search. Therefore, the rate constants k_1, k_2, and pool size of the labile C-pool C_1 were changed stepwise to calculate the parameter response surface functions using approach 2 (*constrained fitting* on *flux*) and 4 (*constrained fitting* on *% TOC mineralized*). The C-pool size of C_2 was calculated by 1–C_1 on a ratio basis (0 to 1). Each parameter was changed 200 times within given bounds resulting in a total of 40,000 parameter combinations. Based on these 40,000 model runs and the knowledge of the parameters used in each single run and the calculated SSR (Eq. [2]), as criteria for the goodness of the fit, the 3D parameter space was plotted.

Setup of literature database

For the analysis of published reported kinetic fits, literature was screened and all relevant data including the kinetic parameters (k_n and C_n) and the way the data were fitted were compiled in a database. The gathered publications were arbitrarily selected and only a small proportion of a large number of available publications was screened. Hereby, literature from the 1960s and 1980s were not selected because kinetic fitting was mainly performed using single-pool models using the analytical approach. A list of all used publications is provided in Tab. 1. In total 504 individual incubation experiments extracted from 40 publications were selected. The 504 entries covered publications from the years 1988 to 2016 analysing soils from around the globe (Africa = 16, Asia = 88, Australia = 16, East Antarctica = 32, Europe = 152, North America = 193, Oceania = 4, and South America = 3) varying in soil texture (sand = 52 %, silt = 20 %, clay = 5 %, no information = 23 %), sampling depth (topsoil = 90 %, subsoil = 7 %, no information = 3 %), and soil amendments. For the soil amendments biochar was the largest percentage followed by plants/plant residues (see Tab. 2). The large number of entries using biochar as an amendment might be explained by the fact that biochar was studied intensively within the last years. Nevertheless, in 45 % of all entries non-amended soils were also reported, which are classically reference soils. The incubation times also vary widely from 0.5 to 800 days, whereby incubation temperatures from 5 to 40°C are reported. Overall, 548 kinetic fits are presented, which is larger than the total number of individual incubation studies, but can be explained by the fact that some incubation results were fitted by different models. From the fitted models the double-pool model was used in 64 % and the single-pool model in only 36 % of all cases to describe the observed data.

Data digitalization

Because in all publications raw data of the incubation experiment (time vs *flux* or time vs *% TOC mineralized*) are not reported in tabular form the presented scatterplots were digitized for further analysis. Therefore, screenshots of the diagrams were imported to the digitalization program *plot digitizer 2.6.8* (Huwaldt & Steinhorst, 2015). First, the origin of the axes as well as the highest x and y-values were defined and the diagram plane was spanned. After, all point values were picked and an output table with the $x - y$ pairs (time vs *flux* or time vs *% TOC mineralized*) was generated and stored.

Results

Effect of different data expressions and fitting constraints on kinetic parameters

As mentioned in the *Background and Theory* section different data expressions (*flux* vs *% TOC mineralized*) and data constraints (*constrained* and *unconstrained*) are reported in literature leading to 4 different fitting approaches, which are possible to fit a kinetic model to the same experimental incubation data.

To analyze the effect of the 4 different approaches on the kinetic parameters obtained by fitting a double-pool model (Eq. [1] with $n = 2$) data published by Muhkerjee et al. (2015) were fitted using the 4 approaches. The data used were measured from a sandy soil amended with 5 % digestate and 1 % high temperature biochar incubated at ~ 20°C for 90 days. Based on the addition of digestate and biochar conceptually more than one C-pool should exist, and therefore, a double-pool model should be appropriate to reproduce the experimental findings. For the *constrained fitting* on *flux* data (approach 1) the sum of the two pools C_1 and C_2 matched the total cumulative flux at t_{end} of 0.0074766 mg C g mixture^{-1}. The model results of the fitting of the 4 different approaches are plotted in Fig. 4a and 4b. As can be seen, all approaches fitted the data well using the double-pool model. On the other hand, looking at the kinetic parameters fitted (Tab. 3) we can see clear differences between the fitting approaches. For the *flux* data the first two fitting approaches (1 and 2) yielded the same fit curves for the total flux as well as same kinetic parameters, which could be expected, because C_1 and C_2 can be expressed in two ways (% flux or mass at t_{end}), while containing the same information content. Therefore, kinetic parameters obtained by both approaches can be directly compared with each other within or between studies. On the other hand, for the fit where C_1 and C_2 are fitted without a constraint (approach 3), the model fit looks slightly different, especially close to the last measured data point. As a consequence, the kinetic parameters differ also from those obtained by approach 1 and 2, whereby both rate constants (k_1 and k_2) are larger resulting in smaller half-lives of 9.1 and 251 days compared to 11.2 and 871 days fitted by approach 1 and 2. Additionally, the C-pool sizes also differ from approach 1 and 2 and the sum of both pools is 0.1159 mg C g mixture^{-1}, which is 2,432% of the observed flux. This problem of overestimation can be also observed in literature, where Zimmerman et al. (2011) for example used such unconstrained fitting (fit on SF33 – in supplementary data associated with the article) and the sum of the two pools estimated by an unconstrained double-pool fit was about 280 % of the total observed flux from the system. Because of the potential problem in overestimating the pool sizes by this approach, the unconstrained fit should be avoided. Finally, the data transformed to *% TOC mineralized* was also fitted and the kinetic parameters show again different values as those obtained from the other approaches. Hereby, the fast pool showed faster turnover of 9.2 days and the slow of 1575 days, which means for the slow pool a 181 % longer turnover. Additionally, the C-pool sizes differ from those fitted by approach 1 and 2, whereby the faster pool in approach 4 is now the smaller pool (18.8 % of the C-stock) compared to approach 1 and 2, where the fast C-pool makes up 97.3 % of the respiration. The general behavior of differences of the kinetic parameters from different fitting approaches has been already observed by Mtambanengwe & Kirchmann (1995) but not further discussed.

The reasons for the reported differences are quite clear. For the fitting on *flux* data the kinetic parameters are only representative for the C-stock contributing to the flux, whereas for the *% TOC mineralized* approach the turnover refers to the entire carbon stock in the soil. In consequence of these differences in the obtained kinetic parameters the extracted information should be interpreted carefully. First, within one study the fitting approaches 1, 2, and 4 are feasible and provide useful information to interpret differences between treatments or sampling locations. Unfortunately, approach 1 and 2 are not useful to inform complex predictive carbon models such as RothC, Candy, Century, or Yasso because the fitted rate constants (k_n) will always refer only to the carbon turned over and not to the total C-stock, and are therefore far too large (or half-lives will be too short). In any case, the exact way how the data were fitted should be described to allow comparison between studies. On the other hand, comparison between kinetic parameters from one study to other reported studies is not straight forward if the procedure of kinetic parameter estimation is not reported. Finally, it has to be noted that the problem described is not inherent to the fitting of double-pool models only but will also occur if a single-pool model will be used.

Over-fitting

Another problem which might occur during fitting of the kinetic model to the incubation data is over-fitting, which is defined as the instance that a simpler model with less parameters can describe the measured data equally well as a more complex model containing more fitting parameters. As an example of over-fitting, we selected the digitized data of Qayyum et al. (2012) for the Oxisol amended with high temperature biochar (HTB) and fitted a single- and double-pool model to the reconstructed measurements using the *constrained* fitting on *flux* data (approach 1). The results are plotted in Fig. 5 and indicate that both models can equally well describe the measured data. Looking at the statistical matrices the SSR (Eq. [2]) and χ^2 (Eq. [3]) values are 328567 and 5.2 for the single-pool fit and 32857 and 5.6 for the double pool model. The same calculated SSR for both fits indicat the same goodness of the two models but the larger χ^2 for the double-pool fit nicely shows that this model is less suitable compared to the single-pool model. If we look closer to the fitted kinetic parameters, we will see that the fitted rate constants k_1 and k_2 for the double-pool model are 0.00794 (HL = 87.3) and 0.00791 (HL = 87.6) days, respectively. Here, it becomes obvious that the two different pools C_1 (= 3698 mg kg^{-1}) and C_2 (= 2522 mg kg^{-1}) turn over with nearly the same speed and can therefore not be disentangled from each other. In consequence, these data should be fitted using a single-pool model only, where the rate constant was estimated to be 0.00792 (HL = 87.5) days.

Ill-posed fitting

A second general problem of fitting a double-pool model to incubation data is associated to the ill-posed problem or non-uniqueness, which will result in unreliable kinetic parameters. To get an impression of the ill-posed problem we calculated the parameter space for the data reported by Mukherjee et al. (2015). In general, the data should contain conceptually at least two pools because the soil respiration was measured from a mixture of a sandy soil with 5 % digestate and 1 % biochar. The bounds for the parameter space were set to $0.01 \geq k_1 \leq 0.15$, $0.0001 \geq k_2 \leq 0.001$ and $0.01 \geq C_1 \leq 0.3$ for approach 1 and to $0.01 \geq k_1 \leq 0.15$, $5x10^{-6} \geq k_2 \leq 1x10^{-4}$ and $0.001 \geq C_1 \leq 0.015$ for approach 2, respectively. For each parameter combination the modeled data were compared to the measured ones and the mismatch between the two was expressed in SSR (Eq. [2]). The corresponding parameter spaces or response functions are given in Fig.6a and b for the fitting using the *constrained* fitting on *flux* (approach 2) and *constrained* fitting on *% TOC mineralized* (approach 4). Additionally, the best fit is indicated by a yellow star within the parameter space. As can be seen, low SSR values are calculated for a wide range of parameter combinations for both data sets, indicating that the topography is fairly flat in the vicinity of the best fit. This means that all those combination in the wide flat area provide nearly the same result in terms of SSR and that all those parameter combinations explain the measured data equally good. To get an impression how well the different parameter combination fit the data, all those combinations showing a difference less than 10 % of the SSR of the optimal fit were selected from the entire data set and the minimum and maximum C_1, k_1 and k_2 combinations were selected. For approach 2 (*constrained* on % *flux*) the estimated size of the slow pool C_1 resulted in a maximum and minimum of 0.69 and 0.66 %, respectively. The rate constant of the slow pool (k_1) also did not vary much between the minimum and maximum with 0.0663 and 0.0691 days resulting in half-lives of 10.0 and 10.5 days, respectively. On the other hand, the slow pool estimates varied between $1.07x10^{-5}$ and $5x10^{-6}$ days resulting in half-lives (HL) from 64,880 to 138,625 days or 177 and 380 years.

For the fit on *% TOC mineralized (*approach 4) the size of the slow pool C_1 varies between 16.7 and 18.8 % (13 % difference). For the rate constant and the corresponding half-lives of the slow pool also only small absolute differences (HL = 7.63 and 9.24 days) were observed. Again, large differences are detectable for the slow pool with HL of 837 and 1.575 days. Irrespectively of these differences in the kinetic parameters for both fitting approaches the fits match quite well the observed data as shown in Fig. 7. This indicates again that the fitting is ill-posed and that multiple solutions will equally well describe the measured data. From the results presented, it becomes clear that the estimation of the rate constant of the slow pool (k_2) is highly uncertain indicating that the incubation experimental data do not contain enough information for the estimation of this parameter. Especially, if the incubation time is short and if the log-transformed data are close to linear, the estimation of the slow pool rate constant (k_2) is critical as already stated by Scharnagl et al. (2010). Based on the results described above, Scharnagl et al. (2010) and Li et al. (2013) proposed to use Bayesian calibration approaches to analyze the real uncertainty and the identifiability of the parameter estimation (fitting). Unfortunately, Bayesian calibration is not as straight forward as classical fitting, which is implemented in most established software packages.

Unrealistic rate constants or half-lives

Based on the two problems described above (over-fitting and ill-posed problem), unrealistic rate constants or half-lives can be estimated. To give an example of unrealistic fitted rate constants we used the kinetic parameters reported by Quayyum et al. (2012) for the control Alfisol subsoil (their original Fig. 2) with k_1 of 3.405 and k_2 of 0.000926 days. In a first step, we plotted the reconstructed double-pool model fit in Fig. 8. In addition, the contributions to the flux from the fast and slow pool were calculated based on the corresponding rate constants (k_n) and pool sizes (C_n). As can be seen, the slow pool flux contributes linear to the bulk flux over the entire duration of the incubation experiment, whereas the fast pool only contributes to the bulk flux within the first day of the experiment as a consequence of the large rate constant k_1 with a resulting half-life of 4.8 hrs. This means, that the fast carbon degraded almost instantaneously, which is biologically not explainable, especially because there has been no fresh and easily degradable carbon source added to the subsoil. It has to be further noted, that in this study, evolved CO_2 was trapped in $KOH_{(aq)}$ and first sampling integrated over the first five days. Therefore, it can be questioned if the five days integration of the flux allows to draw back information about any kinetics shorter as the first sampling.

Exploring the database for fitting problems

In a first step, all database entries were analyzed on the fitting approach used. Unfortunately, only 265 entries out of the 351 reported a double-pool fit could be analyzed because in some cases the entries could not be straight forward reanalyzed due to missing information how the fitting was performed or differences between the equation shown and the plotted data. Out of the total 266 entries the minority (n = 40 or 15 %) was analyzed on converted data to *% TOC mineralized* (approach 4) published in only three publications (Davenport et al., 1988; Collins et al., 2000; Mukherjee et al., 2015). Even if this approach can be used to inform complex predictive carbon models such as RothC, Candy, Century, or Yasso, none of these studies aim to inform such models and the kinetic parameters were only used by Davenport et al. (1988) and Collins et al. (2000) for inter-comparison between studies. For the remaining studies the total evolved cumulative flux at the end of the experiment (t_{end}) was compared to the sum of the estimated C-pools ($C_1 + C_2$). In case that the flux at t_{end} is much smaller as $C_1 + C_2$ it can be assumed that the *unconstrained fitting* on *flux* data (approach 3) was used. Unfortunately, the majority of the entries fall into this approach with 130 entries (49%) from a total of 12 papers (Ajwa & Tabatabi, 1994; Fernández et al., 1999; Devêvre & Horwáth, 2000; Fernández et al., 2001; Pendall & King, 2007; Xuejun et al., 2008; Juarez et al., 2013; Guo et al., 2014; Zhao et al., 2015), whereby, as discussed, this approach will lead to unrealistic C-pool sizes and rate constants. Additionally, seven out of the 12 papers related the estimated parameters to parameters reported by other studies. In comparison, approach 2 (*constrained fitting* on *flux* data using pool sizes expressed in % from total flux) was used 35 times (13%) in four papers (Paul et al., 1999; Cheng et al., 2008; Gillis & Price, 2011; Ci et al., 2015) and approach 1 (*constrained fitting* on *flux* data) was used 60 time (23 %) reported in three publications (Bernal et al., 1998; Haddix et al., 2011; Calvelo Pereira et al., 2014). In conclusion nearly half of the entries seem to use a fitting approach (approach 3), which is not recommended to estimate reliable rate constants and especially C-pool sizes.

In a next step, the data base was analyzed according to the flowchart depicted in Fig. 9. First, obviously wrong reported kinetic parameters were identified and classified into *Class A*. These are parameters of a double-pool model, whereby one estimated k_n or C_n equal zero, which means that either one C-pool does not turn-over ($k_n = 0$) or that the pool turns over but has no C-pool size ($C_n = 0$). In both cases no flux will be formed from the corresponding pool, which will consequently result in a single pool model, and therefore, no double pool model should have been fitted to the data. From all double-pool model data entries (*n* = 336) four were in *Class A*, which corresponds to 1.2 % of the data entries. These four fits originated from two different publications (Cayuela et al., 2010; Zimmermann et al., 2011) and three of them are from soils without amendment.

For the identification of those entries which fall into *Class B*, defined as those with conspicuous large rate constants or small half-lives, all rate constants of the fast C-pool (k_1) were plotted in Fig. 10 (*n* = 336). As can be seen in the histogram, there are two distinct peaks in the reported half-lives for the fast pool – one at half-lives of 20 to 30 days and another for very short half-lives <5 days, but also long half-lives e.g. of up to 347 days are reported in Cheng et al. (2011). The peak at 20 to 30 days nicely coincidence with the half-lives for the fast C-pool defined in complex predictive carbon models (e.g., RothC = 25.3 days, Century = 20-50 days, and Yasso ~ 42 days). Half-lives between 20 and 40 days are reported in eight papers, whereby only four analyzed the data according to approach 4 (*% TOC mineralized*) (Davenport et al., 1988; Collins et al., 2000; Fang et al., 2014; Mukherjee et al., 2015), which should be the choice to get reliable estimates of the half-lives to inform complex predictive carbon models. On the other hand, the extreme short half-lives (minimum reported = 0.18 days by Mukherjee et al. (2015)) seem to be rather unrealistic. If a threshold below the shortest half-lives used in complex predictive carbon models (20 days) will be set, 281 entries (85%) will fall into this class. Because carbon turnover can be much quicker in incubation studies as compared to natural environmental conditions, faster turnover might be also observed. Nevertheless, setting a threshold at very short half-lives of 1 day will still yield 43 entries (13 %) reported in nine papers (Bernal et al., 1998; Fernández et al., 2006; Pendall & King, 2007; Juarez et al., 2012; Quayyum et al., 2012; Guo et l., 2014; Saviozzi et al., 2014; Mukherjee et al., 2015; Zhao et al., 2015), whereby this extremely short half-live seem rather unrealistic, and therefore, these entries were classified into *Class B*. The reasons for the fitting of such short half-lives can be ascribed to the problem of over-fitting or the ill-posed problem.

To get information about entries falling into the third class (*Class C*) a single- and a double-pool model were refitted to all digitized data and the χ^2value was calculated for both fits. The fitting was performed using approach 1 if *flux* data were reported or according to approach 4 if *% TOC mineralized* was reported. Additionally, *Class C* was subdivided into *Class C-1* if a good kinetic fit was reported and if the reported kinetic parameters can be used to match the reconstructed observations. *Class C-2* was defined in a ways that the reported kinetic parameters cannot be used to match the reconstructed observations but a good double-pool model fit can be generally obtained. In the last class (*Class C-3*) refitted entries were allocated if the data were equally good fitted by a single-pool model, which was judged by a smaller χ^2 value for the single-pool model compared to the double-pool model (over-fitted). Unfortunately, for 92 entries the fitting approach could not be clearly identified, and therefore, these entries were not used further. Therefore, in total only 173 of the entries of the entire database could be used for this analysis. In total, the majority of the entries (123 or 70 %) were classified into *Class C-1*, which shows that only less than a third of the analyzed entries (29 %) are over-fitted or that the reported parameters are not the best parameters to describe the experimental data. From this entries 19 (11 %) fall into *Class C-2*, where the kinetic parameters observed do not match the reconstructed measured data but a double-pool can be in general successfully fitted to the observed data. This finding is somehow in contradiction to the statement given in literature where only 5.6 % of all fits where judged to be no good fits (e.g., indicated by low R^2 values) (e.g., in Côté et al., 2000; Guo et al., 2014; Zimmermann et al., 2011), which indicates that most authors assumed that their fits represent the data well and that the estimated kinetic parameters are reliable. On the other hand, *Class C3* includes all over-fitted entries, where a single-pool model should have been fitted to the experimental data. This class covers 32 entries or 18 % from eight individual papers (Davenport et al., 1988; Bernal et al., 1998; Fernández et al., 1999, 2006; Devêvre & Horwáth, 2000; Zimmermann et al., 2011; Juarez et al., 2013; Mukherjee et al., 2015). This clearly indicates that the problem of over-fitting is not well recognized by some scientists even if this problem can be easily overcome by using appropriate statistics and fitting sequentially simple to more complex models to the data.

Discussion and Conclusion

The estimation of kinetic parameters is not straight forward and as shown, different fitting approaches will lead to differences in the estimated kinetic parameters, over-fitting has to be avoided, and the ill-posed problem has to be tackled. Therefore, alternatives to the fitting where the kinetic parameters are not essentially needed are required. One widely used possibility is to use the cumulated fluxes at t_{end} to compare different treatments or locations within one study. This approach is already widely applied and used by e.g. Marschner & Noble (2000), Bruun et al. (2011) amongst many others, but does not allow comparison between studies nor the parameterization of complex predictive carbon models.

If kinetic fitting will be performed, the authors should first define the aim of the study. If only different treatments within one study will be compared, or if they aim for inter-study comparison the fitting approach 1, 2, or 4 should be selected. In any case, care should be taken to compare own results only with those reported results using the same fitting approach.

In any case the unconstrained fitting (approach 3) should be avoided. Our analysis also showed that inter-comparison of estimated parameters between different incubation studies has only been used in half of the studies and the potential of parameter inter-comparison has therefore not yet been fully explored. Again, to provide a meaningful analysis of differences such inter-comparisons require a detailed description of the applied approaches. If the aim of the study would be the information of complex predictive carbon models (e.g., RothC, Century, Candy, Yasso) only the fitting approach 4 should be selected to generate the rate constants or half-lives. Unfortunately, even this approach has a drawback because the experimentally determined (often small) CO_2 flux has to be related to a highly uncertain and often magnitude larger TOC stock.

During the fitting exercise, irrespectively which approach selected, the methodology provided in the Materials and Methods section should be followed in any case, which included the application of an appropriate statistical metric for the judgment of over-fitting, plotting the model fit along with the measured data, analyzing the residues and a careful check of the individual C-pool contributions. Especially the plotting of the model results along with the measured data will allow the reader to judge the goodness of the fit. Unfortunately, in some cases spline curves were included instead in the plots of measured data (e.g., Ajwa & Tabatabai, 1994 (their Fig. 1-3), Mtambanengwe & Kirchmann, 1995 (their Fig. 1), Qayyum et al., 2011 (their Fig. 2), Ouyang et al., 2008 (their Fig. 1), Zimmermann et al., 2011 (their Fig 1), Calvelo Pereira et al., 2014 (their Fig. 5), Devêvre & Horwáth. 2000 (their Fig. 1), Pendall & King, 2007 (their Fig. 1 and 2), Tian et al., 2011 (their Fig. 1), Guo et al., 2014 (their Fig. 2), Ci et al., 2015 (their Fig. 1)) which might lead to misinterpretation of the goodness of the fit.

All these recommendations will not overcome the ill-posed problem, which is inherent to the experimental data and the nature of the simple additive carbon decay (double-pool) model given in Eq. [1]. As an alternative to the double-pool model a single-pool model can be fitted to the data (model choice depends on goodness of fit) and the corresponding half-lives (HL) or mean residence times (MRT) can be calculated for the bulk *flux* (or *% TOC mineralized*). If the single-pool model cannot describe the data sufficiently good a double-pool model can be also fitted but instead of providing the pool sizes and rate constants for both pools a bulk half-life can be estimated numerically. From the knowledge of the bulk HL or MRT in the treatments with respect to a control, the changes in bulk HL or MRT can be easily calculated and attributed to the treatment. The change in bulk HL or MRT between treatment and reference (especially if additional C sources will be added such as biochar or compost) will provide also direct information about the turnover of the amendment.

Finally, it can be questioned if the simple additive carbon model (Eq. [1]) is the right choice to get insight into the complex carbon turnover in soils. In general, in this model the two carbon pools are not interlinked with each other compared to more complex models such as RothC, Candy, Century, or Yasso, which might lead to ill-posed problem. Unfortunately, only little attempt has been made to fit complex models to incubation data. Motavalli et al. (1994) for example used the Century model to estimate the carbon pool sizes and rate constants from long-term (341 days) incubation data and found distinct differences between the rate constants fitted using the classical additive model (Eq. [4]) and Century. This general problem was already observed by Berg and Ågren (1984) who stated, that "the subdivision of the decomposing substrate into chemical constituents can be done in a number of different ways, none of which can be pointed out as the best, but which give rise to very different kinetics." Motavalli et al. (1994) also showed that the regression between the Century and classical derived pool sizes was weak for a wide range of tropical soils. Scharnagl et al. (2010) fitted the RothC model to synthetically generated incubation data and found that the pools can be disentangles at long incubation times (900 days) but still the uncertainty remained quite large. Irrespectively of the advantages of this approach strong model assumptions have to be made controlling carbon flow between the pools and CO_2 release from the soil as stated by Motavalli et al. (1994).

As it is known that conceptual pools such as those defined in the additive model (Eq. [1]) or those integrated into most complex predictive carbon models do not exist in nature, model alternatives avoiding such pool concepts might be favorable. Amongst the first introducing a carbon turnover model without conceptual carbon pools were Ågren & Bosatta (1987) with their carbon quality model (*Q-model*). In this model the bulk organic carbon will not be separated into conceptual pools with assigned rate constants and the respiration flux will be described by one pool (total soil carbon) decaying with a continuous distribution of turnover rates describing different carbon qualities within the bulk C-stock. A comparable model concept has been also introduced by Forney & Rothman (2012 & 2014). Even if the *Q-model* has been tested successfully and has been used to describe incubation data, its general application is still limited, mainly due to its mathematical complexity.

Finally, we have to question if the classical incubation experiments are still state-of-the-art if we want to gain knowledge about the carbon pools and the rate constants, because the use of natural isotopic signatures of the soils and amendments (e.g., Torn et al., 2013) or the isotopic labeling of the amendment (e.g., Séquaris et al., 2010) can nowadays easily be performed. This labeling and signatures can be used to disentangle the C-pools from each other or to study the turnover of an individual pool. Additionally, the increase of the duration of the incubation experiment might also increase the information content of the data, especially for the estimation of the slow pool, as shown for a multi-pool model by Scharnagl et al. (2010).

References

Ågren, G.I. & Bosatta, E. 1987. Theoretical-analysis of the long-term dynamics of carbon and nitrogen in soils. Ecology 68 (5): 1181-1189.

Ajwa, H.A. & Tabatabai M.A. 1994. Decomposition of different organic materials in soils. Biology and Fertility of Soils 18 (3): 175-182.

Akaike, H. 1974. New look at statistical-model identification. IEEE Trans. Autom. Control AC19(6), 716–723.

Barrett, J.E., Virginia, R.A, Parsons, A.N. & Wall, D.H. 2006. Soil carbon turnover in the McMurdo Dry Valleys, Antarctica. Soil Biology and Biochemistry 38 (10): 3065–3082.

Bauer, J.; Weihermüller, L.; Huisman, J.A.; Herbst, M.; Graf, A.; Séquaris, J.-M. & Vereecken, H. 2012. Inverse determination of heterotrophic soil respiration response to temperature and water content under field conditions. Biogeochemistry 108 (1-3): 119-134.

Berg, B., & Agren, G.I. 1984. Decomposition of needle litter and its organic chemical components: theory and field experiments. Long-term decomposition in a Scots pine forest. III. Canadian Journal of Botany 62: 2880-2888.

Bernal, M.P., Sánchez-Monedero, M.A., Paredes, C. & Roig, A. 1998. Carbon mineralization from organic wastes at different composting stages during their incubation with soil. Agriculture, Ecosystems and Environment 69 (3): 175–189.

Beven K. 1993. Prophesy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16: 41-51.

Beven K. 1996. Equifinality and uncertainty in geomorphological modelling. In: The Scientific Nature of Geomorphology. Rhoads B.L. and Thorn C.E. (eds). Wiley: Chichester, UK. 289-313.

Bruun, E.W., Müller-Stöver, D, Ambus, P. & Hauggaard-Nielsen, H. 2011. Application of biochar to soil and N2O emissions: potential effects of blending fast-pyrolysis biochar with anaerobically digested slurry. European Journal of Soil Science 62: 581–589.

Bustamante, M.A., Pérez-Murcia, M.D., Paredes, C., Moral, R., Pérez-Espinosa, A. & Moreno-Caselles, J. 2007. Short-term carbon and nitrogen mineralisation in soil amended with winery and distillery organic wastes. Bioresource Technology 98 (17): 3269–3277.

Calvelo Pereira, R., Camps Arbestain, M., Kaal, J., Vazquez Sueiro, M., Sevilla, M. & Hindmarsh, J. 2014. Detailed carbon chemistry in charcoals from pre-European Māori gardens of New Zealand as a tool for understanding biochar stability in soils. European Journal of Soil Science 65 (January): 83–95.

Cayuela, M.L., Oenema, O., Kuikman, P.J., Bakker, R.R. & Van Groenigen, J.W. 2010. Bioenergy by-products as soil amendments? Implications for carbon sequestration and greenhouse gas emissions. Global Change Biology Bioenergy 2: 201–213.

Cheng, C.-H., Lehmann, J., Thies, J.E. & Burton, S.D. 2008. Stability of black carbon in soils across a climatic gradient. Journal of Geophysical Research 113 (G2): G02027.

Ci, E., Al-Kaisi, M.M., Wang, L., Ding, C. & Xie, D. 2015. Soil Organic Carbon Mineralization as Affected by Cyclical Temperature Fluctuations in a Karst Region of Southwestern China. Pedosphere 25 (4): 512–523.

Coleman, K. & Jenkinson, D.S. 1996. RothC-26-3. A model for the turnover of carbon on soil. Vol. NATO ASI Series I. 38, Evaluation of soil organic matter models using existing, long-term datasets. Springer-Verlag, Berlin.

Coleman, K. & Jenkinson, D.S. 1999. ROTHC-26.3: A model for the turnover of carbon in soil, Model description and windows users guide, IACR, Rothamsted, Harpenden, online available at: http://www.rothamsted.bbsrc.ac.uk/aen/carbon/mod26_3_win.pdf.

Collins, H.P., Elliott, E.T., Paustian, K., Bundy, L.G., Dick, W. A., Huggins, D.R., Smucker, A.J.M. & Paul, E.A. 2000. Soil carbon pools and fluxes in long-term Corn Belt agroecosystems. Soil Biology and Biochemistry 32 (2): 157–168.

Cook, B.D. & Allan, D.L. 1992. Dissolved organic carbon in old field soils: total amounts as a measure of available resources for soil mineralization. Soil Biology and Biochemistry 24 (6): 585–594.

Côté, L., Brown, S., Paré, D., Fyles, J. & Bauhus, J. 2000. Dynamics of carbon and nitrogen mineralization in relation to stand type, stand age and soil texture in the boreal mixedwood. Soil Biology and Biochemistry 32 (8-9): 1079–1090.

Davenport, J.R., Thomas, R.L. & Mott, S.C. 1988. Carbon mineralization of corn (Zea mays L.) and bromegrass (Bromus inermis Leyss.) components with an emphasis on the below-ground carbon. Soil Biology and Biochemistry 20 (4): 471–476.

Devêvre, O.C. and Horwáth, W.R. 2000. Decomposition of rice straw and microbial carbon use efficiency under different soil temperatures and moistures. Soil Biology and Biochemistry 32: 1773–1785.

Duan, Q.Y., Sorooshian, S., & Gupta, V. 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resources Research 28 (4): 1015-1031.

Duan, Q.Y., Sorooshian, S., & Gupta, V. 1994. Optimal use of the SCE-UA global optimization method for calibrating watershed models. Journal of Hydrology 158 (3-4): 265-284.

Fang, Y., Singh, B., Singh, B. P. & Krull, E. 2014. Biochar carbon stability in four contrasting soils. European Journal of Soil Science 65 (1): 60–71.

Fernández, I., Cabaneiro, A. & Carballas, T. 1999. Carbon mineralization dynamics in soils after wildfires in two Galician forests. Soil Biology and Biochemistry 31 (13): 1853–1865.

Fernández, J.M., Plaza, C., Hernández, D. & Polo, A. 2007 Carbon mineralization in an arid soil amended with thermally-dried and composted sewage sludges. Geoderma 137: 497–503.

FOCUS 2006. Guidance Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies on Pesticides in EU Registration, Report of the FOCUS Work Group on Degradation Kinetics, EC Document Reference Sanco/10058/2005 version 2.0, 434 pp.

Forney, D.C. & Rothman, D.H. 2012. Inverse method for estimating respiration rates from decay time series. Biogeosciences 9: 3601-3612.

Forney, D.C. & Rothman, D.H. 2014. Carbon transit through degradation networks. Ecological Monographs 81 (1): 109-129.

Franko, U., Crocker, G., Grace, P., Klir, J., Körschens, M., Poulton, P., & Richter, D. 1997. Simulating trends in soil organic carbon in long-term experiments using CANDY model. Geoderma 81, 109{120.

Gillis, J.D. & Price, G.W. 2011. Comparison of a novel model to three conventional models describing carbon mineralization from soil amended with organic residues. Geoderma 160 (3-4): 304–310.

Guo, J., Yang, Y., Chen, G., Xie, J. & Yang, Z. 2014. Carbon mineralization of Chinese fir (Cunninghamia lanceolata) soils under different temperature and humidity conditions.

Acta Ecologica Sinica 34 (1): 66–71.

Haddix, M.L., Plante, A.F., Conant, R.T., Six, J., Steinweg, J.M., Magrini-Bair, K., Drijber, R.A., Morris, S.J. & Paul, E.A. 2011. The Role of Soil Characteristics on Temperature Sensitivity of Soil Organic Matter. Soil Science Society of America Journal 75 (1): 56–68.

Hansen, S., Jensen, H., Nielsen, N., & Svendsen, H. 1990. DAISY - Soil plant atmosphere system model. Tech. rep., The National Agency for Environmental Protection.

Hess, T.F. & Schmidt, S.K. 1995. Improved procedure for obtaining statistically valid parameter estimates from soil respiration data. Soil Biology and Biochemistry 27: 1–7.

Hopkins, D.W,, Sparrow, A.D., Elberling, B., Gregorich, E.G., Novis, P.M., Greenfield, L.G. & Tilston, E.L. 2006. Carbon, nitrogen and temperature controls on microbial activity in soils from an Antarctic dry valley. Soil Biology and Biochemistry 38 (10): 3130–3140.

Hurvich, C.M. & Tsai, C.L. 1989. Regression and time-series model selection in small samples. Biometrika 76, 297–307.

Huwaldt, J.A. & Steinhorst, S. 2015. Plot Digitizer version 2.6.8, accessed October 27, 2015 at http://plotdigitizer.sourceforge.net/.

Jha, P., Garg, N., Lakaria, B.L., Biswas, A.K. & Rao, A.S. 2012. Soil and residue carbon mineralization as affected by soil aggregate size. Soil and Tillage Research 121: 57–62.

Jia, J., Yu, D., Zhou, W., Zhou, L., Bao, Y., Meng, Y. & Dai, L. 2015. Variations of soil aggregates and soil organic carbon mineralization across forest types on the northern slope of Changbai Mountain. Acta Ecologica Sinica 35 (2): 1–7.

Juarez, S., Nunan, N., Duday, A.-C., Pouteau, V., Schmidt, S., Hapca, S., Falconer, R., Otten, W. & Chenu, C. 2013. Effects of different soil structures on the decomposition of native and added organic carbon. European Journal of Soil Biology 58: 81–90.

Li, D., Schädel, C., Haddix, M.L., Paul, E.A., Conant, R., Li, J., Zhou, J. & Luo, Y. 2013. Differential responses of soil organic carbon fractions to warming: Results from an analysis with data assimilation. Soil Biology and Biochemistry 67: 24 – 30.

Liang, B., Lehmann, J., Solomon, D., Sohi, S., Thies, J.E., Skjemstad, J.O., Luizão, F.J., Engelhard, M.H., Neves, E.G. & Wirick, S. 2008. Stability of biomass-derived black carbon in soils. Geochimica et Cosmochimica Acta. 72 (24): 6069-6078.

Liski, J., Palosuo, T., Peltoniemi, M. & Sievänen, R. 2005. Carbon and decomposition model Yasso for forest soils. Ecological Modelling 189(1-2): 168-182.

Liski, J., Nissinen, A., Erhard, M. & Taskinen, O. 2003. Climatic effects on litter decomposition from arctic tundra to tropical rainforest. Global Change Biology 9(4): 575-584.

Marchetti, R., Lazzeri, L., D’Avino, L. & Ponzoni, G. 2015. Nitrogen and carbon mineralization in soils amended with biofumigant or non-biofumigant plant materials. Industrial Crops and Products 75: 65–72.

Marinari, S., Lagomarsino, A., Moscatelli, M.C., Di Tizio, A. & Campiglia, E. 2010. Soil carbon and nitrogen mineralization kinetics in organic and conventional three-year cropping systems. Soil and Tillage Research 109 (2): 161–168.

Marschner, B. & Noble, A.D. 2000. Chemical and biological processes leading to the neutralisation of acidity in soil incubated with litter materials. Soil Biology and Biochemistry 32: 805-813.

Mboh, C.M., Huisman, J.A. & Vereecken, H. 2011. Feasibility of Sequential and Coupled Inversion of Time Domain Reflectometry Data to Infer Soil Hydraulic Parameters under Falling Head Infiltration. Soil Science Society of America Journal 75 (3): 775-786.

Mertens, J.; Madsen, H.; Kristensen, M.; Jacques, D. & Feyen, J. 2005. Sensitivity of soil parameters in unsaturated zone modelling and the relation between effective, laboratory and in situ estimates. Hydrological Processes 19 (8): 1611-1633.

Mohanty, S., Nayak, A.K., Kumar, A., Tripathi, R., Shahid, M., Bhattacharyya, P., Raja, R. & Panda, B.B. 2013. Carbon and nitrogen mineralization kinetics in soil of rice-rice system under long term application of chemical fertilizers and farmyard manure. European Journal of Soil Biology 58: 113–121.

Moreno-Cornejo, J., Zornoza, R. & Faz, A. 2014. Carbon and nitrogen mineralization during decomposition of crop residues in a calcareous soil. Geoderma 230-231: 58–63.

Motavalli, P.P., Palm, C.A., Parton, W.J., Elliott, E.T. & Frey S.D. 1994. Comparison of laboratory and modeling simulation methods for estimating soil carbon pools in tropical forest soils. Soil Biology and Biochemistry 26 (8): 935-944.

Mtambanengwe, F. & Kirchmann, H. 1995. Litter from a tropical savanna woodland (miombo): chemical composition and C and N mineralization. Soil Biology & Biochemistry 27 (12): 1639–1651.

Mukherjee, S., Weihermüller, L., Tappe, W., Vereecken, H. & Burauel, P. 2015. Microbial respiration of biochar- and digestate-based mixtures. Biology and Fertility of Soils 52 (2): 151-164.

Nash, J.E. &Sutcliffe, J.V. 1970. River flow forecasting through conceptual models part I – a discussion of principles. Journal of Hydrology 10, 282–290.

Olsen, J.S. 1963. Energy storage and the balance of producers and decomposers in ecological systems. Ecology 44: 322 - 331.

Ouyang, X., Zhou, G., Huang, Z., Zhou, C., Li, J., Shi, J. & Zhang, D. 2008. Effect of N and P addition on soil organic C potential mineralization in forest soils in South China. Journal of Environmental Sciences (China) 20 (9): 1082–1089.

Parton, W., Pulliam, W. & Ojima, D. 1994. Application of the CENTURY model across the LTR network: Parameterization and climate change simulations. Bulletin of the Ecological Society of America 75, 186{187.

Paul, E.A., Harris, D., Collins, H.P., Schulthess, U. & Robertson, G.P. 1999. Evolution of CO_2 and soil carbon dynamics in biologically managed, row-crop agroecosystems. Applied Soil Ecology 11 (1): 53–65.

Pedra, F., Polo, A., Ribeiro, A. & Domingues, H. 2007. Effects of municipal solid waste compost and sewage sludge on mineralization of soil organic matter. Soil Biology and Biochemistry 39: 1375–1382.

Pendall, E. & King, J.Y. 2007. Soil organic matter dynamics in grassland soils under elevated CO_2: Insights from long-term incubations and stable isotopes. Soil Biology and Biochemistry 39 (10): 2628–2639.

Qayyum, M.F., Steffens, D., Reisenauer, H.P. & Schubert, S. 2011. Kinetics of Carbon Mineralization of Biochars Compared with Wheat Straw in Three Soils. Journal of Environment Quality 41 (4): 1210-1220.

Redin, M., Recous, S., Aita, C., Dietrich, G., Skolaude, A.C., Ludke, W.H., Schmatz, R. & Giacomini, S.J. 2014 How the chemical composition and heterogeneity of crop residue mixtures decomposing at the soil surface affects C and N mineralization. Soil Biology and Biochemistry 78: 65–75.

Reiners, W.A. 1968. Carbon dioxide evolution from the floor of three Minnesota forests. Ecology 49: 471 – 483.

Saviozzi, A., Levi-Minzi, R., Riffaldi, R. & Vanni, G. 1997. Role of chemical constituents of wheat straw and pig slurry on their decomposition in soil. Biology and Fertility of Soils 25 (4): 401-406.

Saviozzi, A., Vanni, G. & Cardelli, R. 2014. Carbon mineralization kinetics in soils under urban environment. Applied Soil Ecology 73: 64–69.

Scharnagl, B., Vrugt, J.A., Vereecken, H. & Herbst, M. 2010 Information content of incubation experiments for inverse estimation of pools in the Rothamsted carbon model: a Bayesian perspective. Biogeosciences 7: 763–776.

Séquaris, J.M., Herbst, M., Weihermüller, L., Bauer, J. & Vereecken, H. 2010. Simulating decomposition of 14C-labelled fresh organic matter in bulk soil and soil particle fractions at various temperatures and moisture contents. European Journal of Soil Science 61: 940–949.

Ŝimũnek, J. & Hopmans, J.W. 2002. Parameter optimization and nonlinear fitting. Dane J.H. & Topp, G.C. (eds.): Methods of soil Analysis, Part 4, physical methods. 3rd edition. Soil Science Society of America, Madison, WI. Chapter 1.7: 139-156.

Spiess, A.-N. & Neumeyer, N. 2010. An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. BMC Pharmacology 10: 6. DOI: 10.1186/1471-2210-10-6

Standford, G. & Smith, S.J. 1972. Nitrogen mineralization potentials of soils. Soil Science Society of America Proceedings 36: 465-472.

Taylor, B.R. & D. Parkinson. 1988. Respiration and mass loss rates of aspen and pine leaf litter decomposing in laboratory microcosms. Canadian Journal of Botany 66: 1948 – 1959.

Tian, Y., Liu, J., Wang, X. & Gao, L. 2011. Carbon mineralization in the soils under different cover crops and residue management in an intensive protected vegetable cultivation. Scientia Horticulturae 127 (3): 198–206.

Torn, M.S., Kleber, M., Zavaleta, E.S., Zhu, B., Field, C.B. & Trumbore, S.E. 2013. A dual isotope approach to isolate soil carbon pools of different turnover times. Biogeosciences 10, 8067–8081.

Trumbore, S.E. 1993. Comparison of carbon dynamics in tropical and temperate soils using radiocarbonmeasurements. Global Biogeochemical Cycles. 7 (2): 275-290.

Weihermüller, L.; Graf, A.; Herbst, M. & Vereecken H. 2013. Simple pedotransfer functions to initialize reactive carbon pools of the RothC model. European Journal of Soil Science 65 (5): 567-575

Weihermüller, L.; Huisman, J.A.; Graf, A.; Herbst, M.; & Séquaris, J.-M. 2009. Multistep Outflow Experiments to Determine Soil Physical and Carbon Dioxide Production Parameters. Vadose Zone Journal 8 (3): 772-782.

Zhao, R., Coles, N. & Wu, J. 2015. Carbon mineralization following additions of fresh and aged biochar to an infertile soil. Catena 125: 183–189.

Zimmerman, A.R., Gao, B. & Ahn, M.Y. 2011. Positive and negative carbon mineralization priming effects among a variety of biochar-amended soils. Soil Biology and Biochemistry 43: 1169 – 1179.

Figure Captions

Figure 1: Examples of the two different ways to express incubation data. A) expression as cumulative CO_2 or CO_2-C equivalent flux and b) as % TOC mineralized.

Figure 2: Synthetically generated incubation flux (or % TOC mineralized) data for a single-pool (a) and double-pool (c) and corresponding log-transformation for the single-pool (b) and double-pool (d) model. Lines in the log transformation indicate linear regressions, whereby for (d) brake point was arbitrarily chosen.

Figure 3: Examples of residue distribution a) for a 'good' single-pool fit and b) a 'bad' single-pool fit.

Figure 4: Fitted model results of the 4 different approaches (1 = *Constrained fitting* on *flux* data using pool sizes expressed on mass basis, 2 = *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux, 3 = *Unconstrained fitting* on *flux* data using pool sizes expressed on mass basis, and 4 = *Constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized). Kinetic parameters are listed in Tab. 3.

Figure 5: Example of over-fitted incubation data. Measured data digitized and reconstructed from Qayyum et al. (2012) (Oxisol + HTC) and fitted single- and double-pool model.

Figure 6: Parameter spaces (response function) for a double pool model Eq. [1] for two different fitting approaches. a) *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux (approach 2) and b) *constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized (approach 4). Yellow star indicates the best parameter combination listed in Tab. 3 and dots indicate selected parameter combinations used in Fig. 7. Measured incubation data taken from Muhkerjee et al. (2015).

Figure 7: Measured a) CO_2 flux and b) % TOC mineralized with best fit and fits based on selected parameter combinations with less than 10% difference from SSR of best fit. Measured data taken from Muhkerjee et al. (2015).

Figure 8: Normalized bulk CO_2 flux and corresponding contributions from the slow and fast pool over time. Fluxes were reconstructed from kinetic parameters reported by Qayyum et al. (2012) for the control Alfisol subsoil (original Fig. 2).

Figure 9: Flowchart for the data base analysis.

Figure 10: Histogram of half-lives [days] of the fast C-pool extracted from the literature data base. Grey bar indicate span of half-lives used in complex predictive carbon models.

Table Captions

Table 1: Overview of the used publications in the database.

Table 2: Overview of the reported amendments in the database.

Table 3: Table 3: Kinetic parameters of the 4 different fitting approaches (1 = *Constrained fitting* on *flux* data using pool sizes expressed on mass basis, 2 = *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux, 3 = *Unconstrained fitting* on *flux* data using pool sizes expressed on mass basis, and 4 = *Constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized). Fitted model results are plotted in Fig. 4. SSR = sum of squared residuals, HL = half-life, and MRT = mean residence time.

Tables

Table 1: Overview of the used publications in the database.

Author	Journal	Year
Ajwa & Tabatabai 1994	Biology and Fertility of soils	1994
Barrett et al. 2006	Soil Biology and Biochemistry	2006
Bernal et al. 1998	Agriculture, Ecosystems & Environment	1998
Bustamante et al. 2007	Bioresource Technology	2007
Calvelo Pereira et al. 2014	European Journal of Soil Science	2014
Cayuela et al. 2010	Global Change Biology Bioenery	2010
Cheng et al. 2008	Journal of Geophysical Research	2008
Ci et al. 2015	Pedosphere	2015
Collins et al. 2000	Soil Science Society of America Journal	2000
Cook & Allan 1992	Soil Biology and Biochemistry	1992
Côté et al. 2000	Soil Biology and Biochemistry	2000
Davenport et al. 1988	Soil Biology and Biochemistry	1988
Devêvre & Horwáth 2000	Soil Biology and Biochemistry	2000
Fang et al. 2014	European Journal of Soil Science	2014
Fernández et al. 1999	Soil Biology and Biochemistry	1999
Fernández et al. 2007	Geoderma	2007
Gillis & Price 2011	Geoderma	2011
Guo et al. 2014	Acta Ecologica Sinica	2014
Haddix et al. 2011	Soil Science Society American Journal	2011
Hopkins et al. 2006	Soil Biology and Biochemistry	2006
Jha et al. 2012	National Acaddemic Science Letters	2012
Jia et al. 2015	Acta Ecologica Sinica	2015
Juarez et al. 2013	European Journal of Soil Biology	2013
Marchetti et al. 2015	Industrial Crops and Products	2015
Marinari et al. 2010	Soil & Tillage Research	2010
Mohanty et al. 2013	European Journal of Soil Biology	2013
Moreno-Cornejo et al. 2014	Geoderma	2014
Mtambanengwe & Kirchmann 1995	Soil Biology and Biochemistry	1995
Mukherjee et al. 2015	Biology and Fertility of soils	2015
Paul et al. 1999	Applied Soil Ecology	1999
Pedra et al. 2007	Soil Biology and Biochemistry	2007
Pendall & King 2007	Soil Biology and Biochemistry	2007
Qayyum et al. 2012	Journal of Environmental Quality	2012
Redin et al. 2014	Soil Biology and Biochemistry	2014
Saviozzi et al. 2014	Applied Soil Ecology	2014
Tian et al. 2011	Scientia Horticilturae	2011
Xuejen et al.2008	Journal of Environmental Sciences	2008
Zhao et al. 2015	Catena	2015
Zimmermann et al. 2011	Soil Biology and Biochemistry	2011

Table 2: Overview of the reported amendments in the database.

Amendment	Total	%
none	225	40
straw	11	2
biochar/charcoal	89	16
manure	30	5
sewage sludge	34	6
N fertilized	18	3
P fertilized	10	2
plants/plant-residues	69	12
compost	4	1
others	72	13

Table 3: Table 3: Kinetic parameters of the 4 different fitting approaches (1 = *Constrained fitting* on *flux* data using pool sizes expressed on mass basis, 2 = *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux, 3 = *Unconstrained fitting* on *flux* data using pool sizes expressed on mass basis, and 4 = *Constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized). Fitted model results are plotted in Fig. 4. SSR = sum of squared residuals, HL = half-life, and MRT = mean residence time.

Approach	Data expression	Constraint	Parameter	Unit	Best fit	SSR	HL [days]	MRT [days]
Approach 1	flux in mass	constraint	C_1	[-]	0.00728	9.14e-7	11.2	16.2
			k_1	[day^{-1}]	0.0618			
			C_2	[-]	0.000199		871	1256
			k_2	[day^{-1}]	0.000796			
Approach 2	flux in %	constraint	C_1	[mg C g mixture^{-1}]	0.00728 (97.3 %)	9.14e-7	11.2	16.2
			k_1	[day^{-1}]	0.0618			
			C_2	[mg C g mixture^{-1}]	0.00199 (2.67 %)		871	1256
			k_2	[day^{-1}]	0.000796			
Approach 3	flux in mass	no constraint	C_1	[mg C g mixture^{-1}]	0.00629	8.10e-7	9.1	13.1
			k_1	[day^{-1}]	0.0761			
			C_2	[mg C g mixture^{-1}]	0.0053		251	362
			k_2	[day^{-1}]	0.00276			
Approach 4	% TOC mineralized	constraint	C_1	[% TOC]	0.188 (18.8 %)	7.005	9.2	13.33
			k_1	[day^{-1}]	0.0750			
			C_2	[% TOC]	0.812 (81.2 %)		1575	2273
			k_2	[day^{-1}]	0.00044			

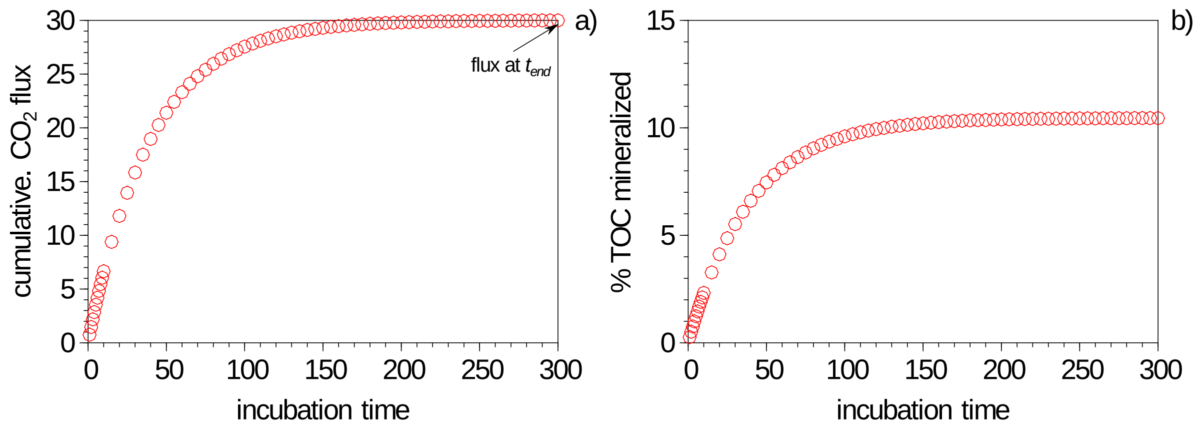

Figure 1: Examples of the two different ways to express incubation data. A) expression as cumulative CO_2 or CO_2-C equivalent flux and b) as % TOC mineralized.

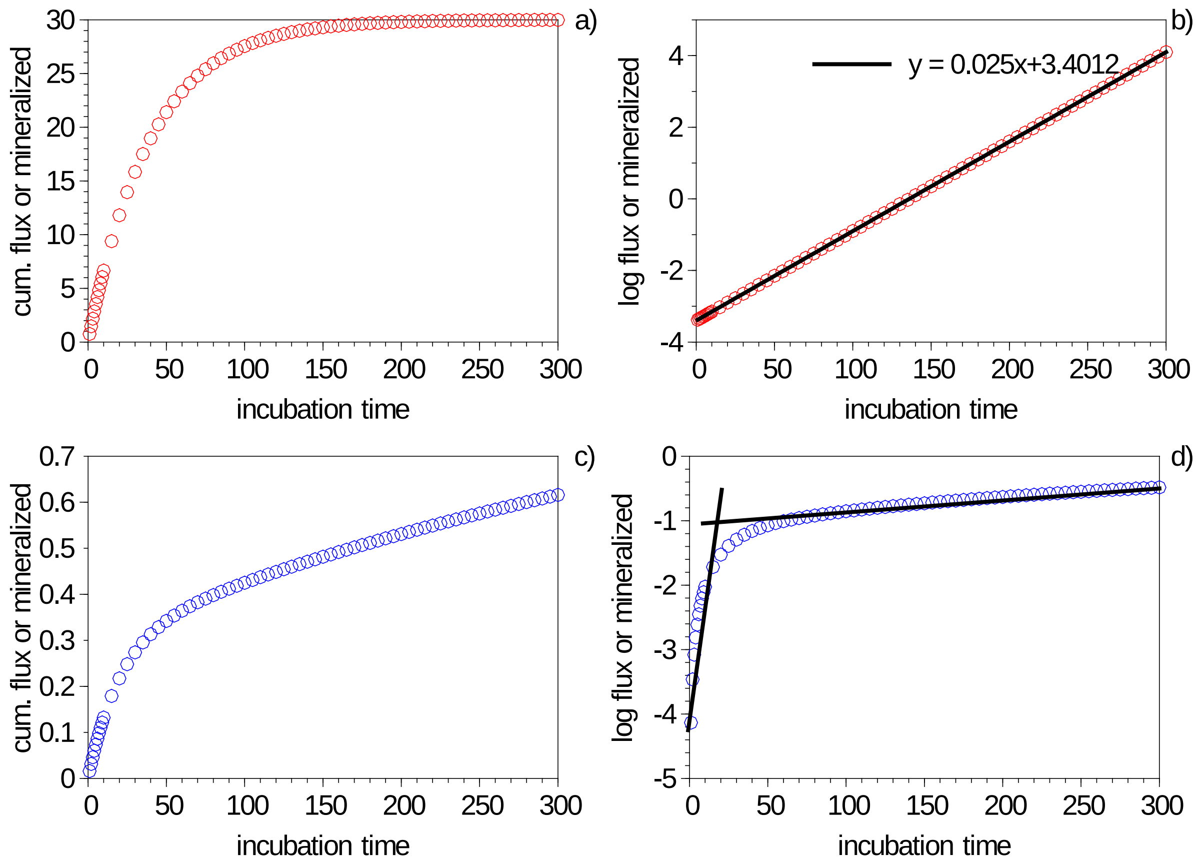

Figure 2: Synthetically generated incubation flux (or % TOC mineralized) data for a single-pool (a) and double-pool (c) and corresponding log-transformation for the single-pool (b) and double-pool (d) model. Lines in the log transformation indicate linear regressions, whereby for (d) brake point was arbitrarily chosen.

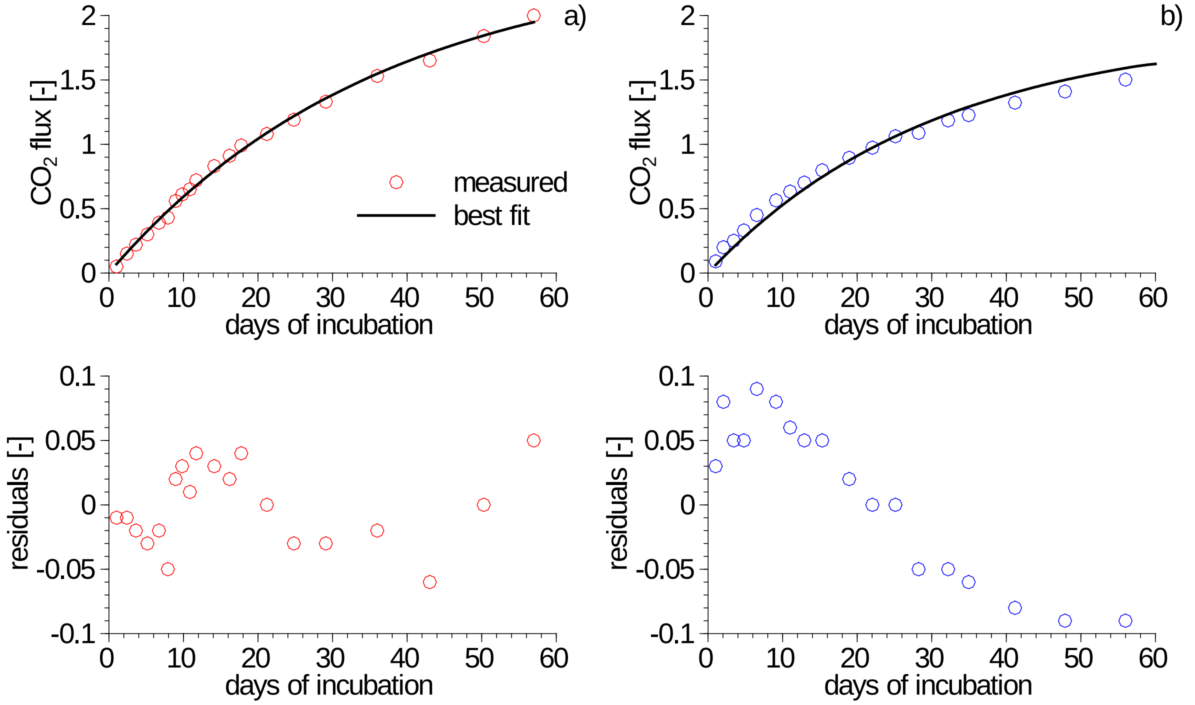

Figure 3: Examples of residue distribution a) for a 'good' single-pool fit and b) a 'bad' single-pool fit.

a)

b)

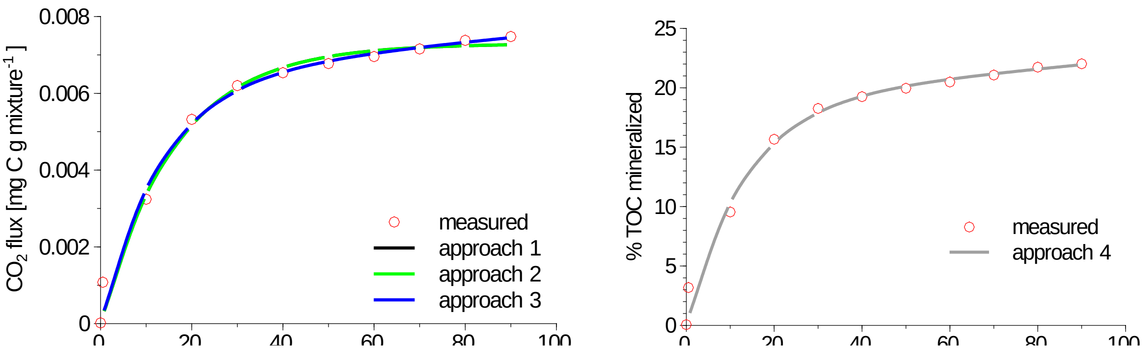

Figure 4: Fitted model results of the 4 different approaches (1 = *Constrained fitting* on *flux* data using pool sizes expressed on mass basis, 2 = *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux, 3 = *Unconstrained fitting* on *flux* data using pool sizes expressed on mass basis, and 4 = *Constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized). Kinetic parameters are listed in Tab. 3.

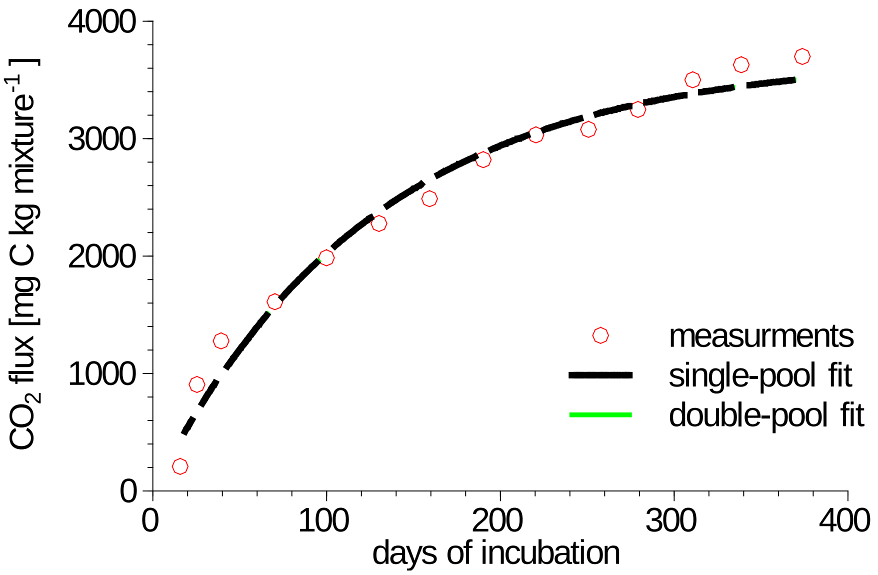

Figure 5: Example of over-fitted incubation data. Measured data digitized and reconstructed from Qayyum et al. (2012) (Oxisol + HTC) and fitted single- and double-pool model.

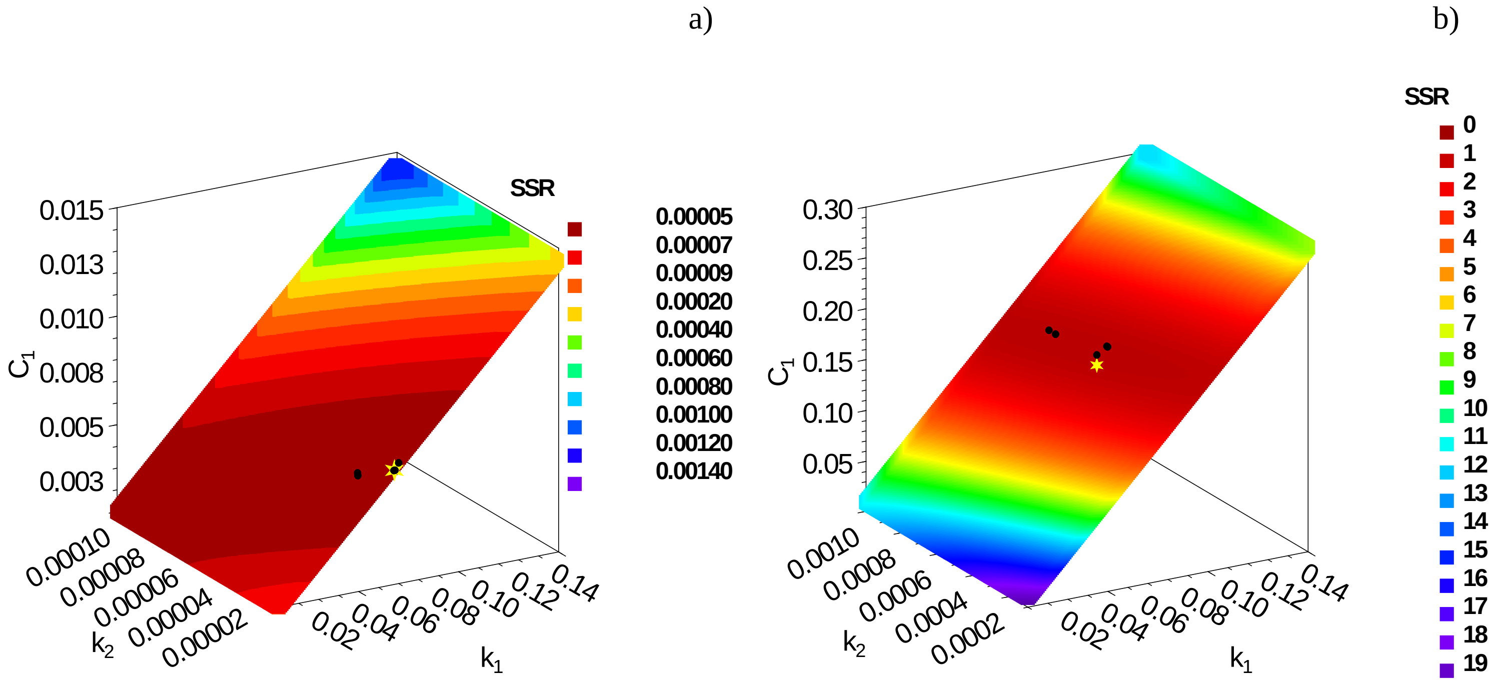

Figure 6: Parameter spaces (response function) for a double pool model Eq. [1] for two different fitting approaches. a) *Constrained fitting* on *flux* data using pool sizes expressed in % from total flux (approach 2) and b) *constrained fitting* on *% TOC mineralized* data using pool sizes expressed in % from total TOC mineralized (approach 4). Yellow star indicates the best parameter combination listed in Tab. 3 and dots indicate selected parameter combinations used in Fig. 7. Measured incubation data taken from Muhkerjee et al. (2015).

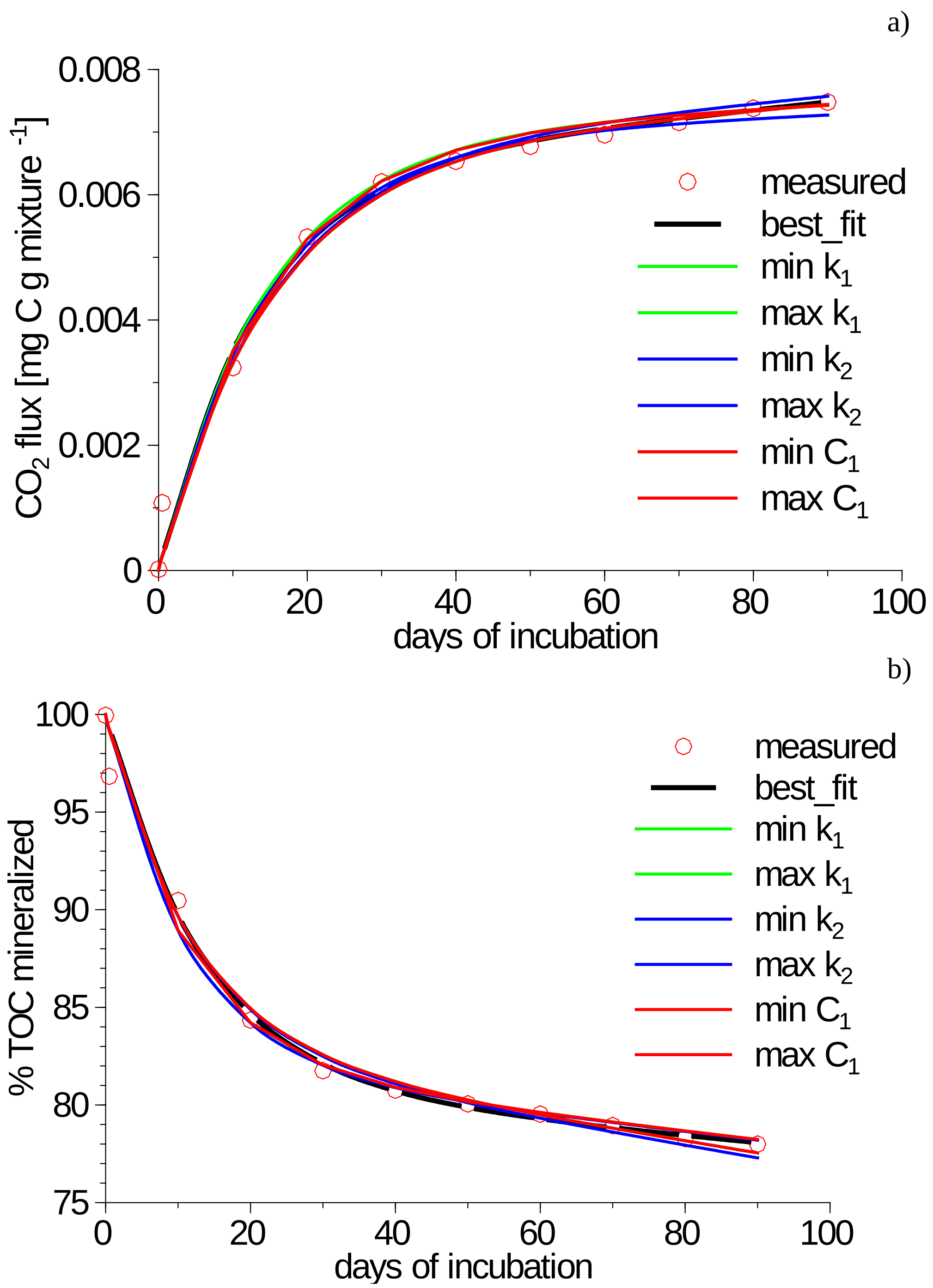

Figure 7: Measured a) CO_2 flux and b) % TOC mineralized with best fit and fits based on selected parameter combinations with less than 10% difference from SSR of best fit. Measured data taken from Muhkerjee et al. (2015).

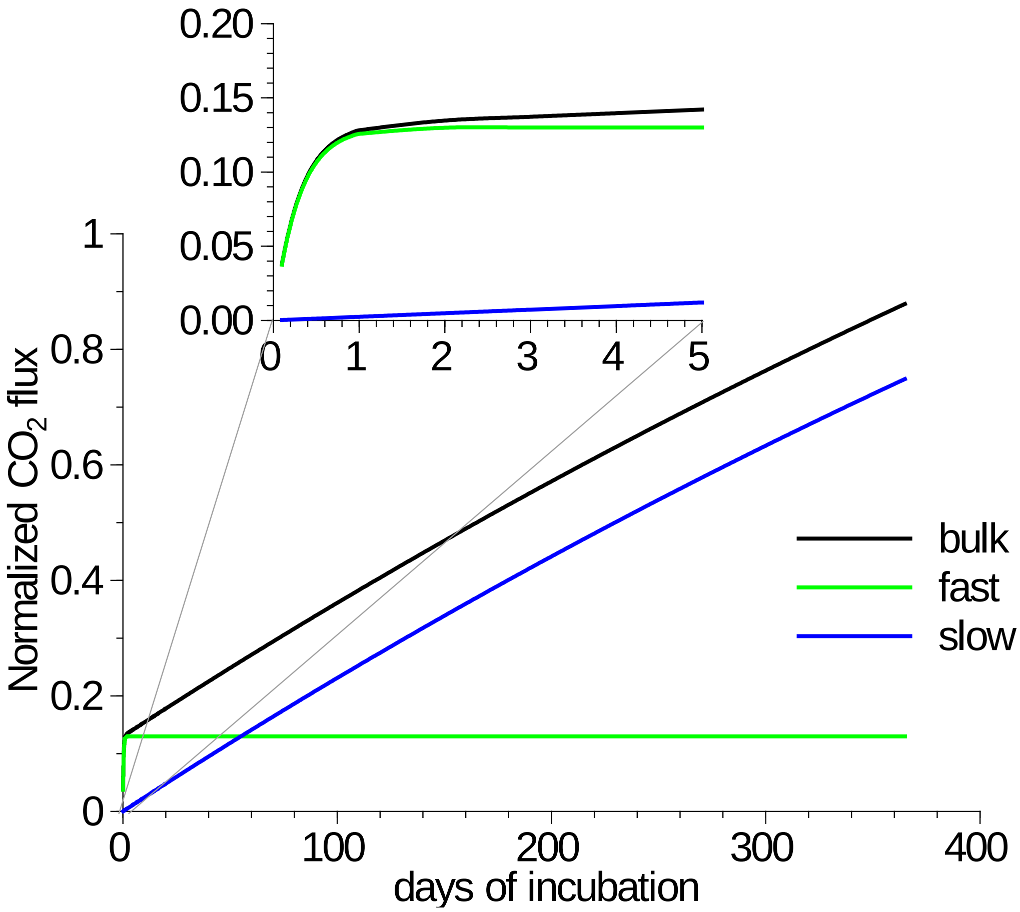

Figure 8: Normalized bulk CO_2 flux and corresponding contributions from the slow and fast pool over time. Fluxes were reconstructed from kinetic parameters reported by Qayyum et al. (2012) for the control Alfisol subsoil (original Fig. 2).

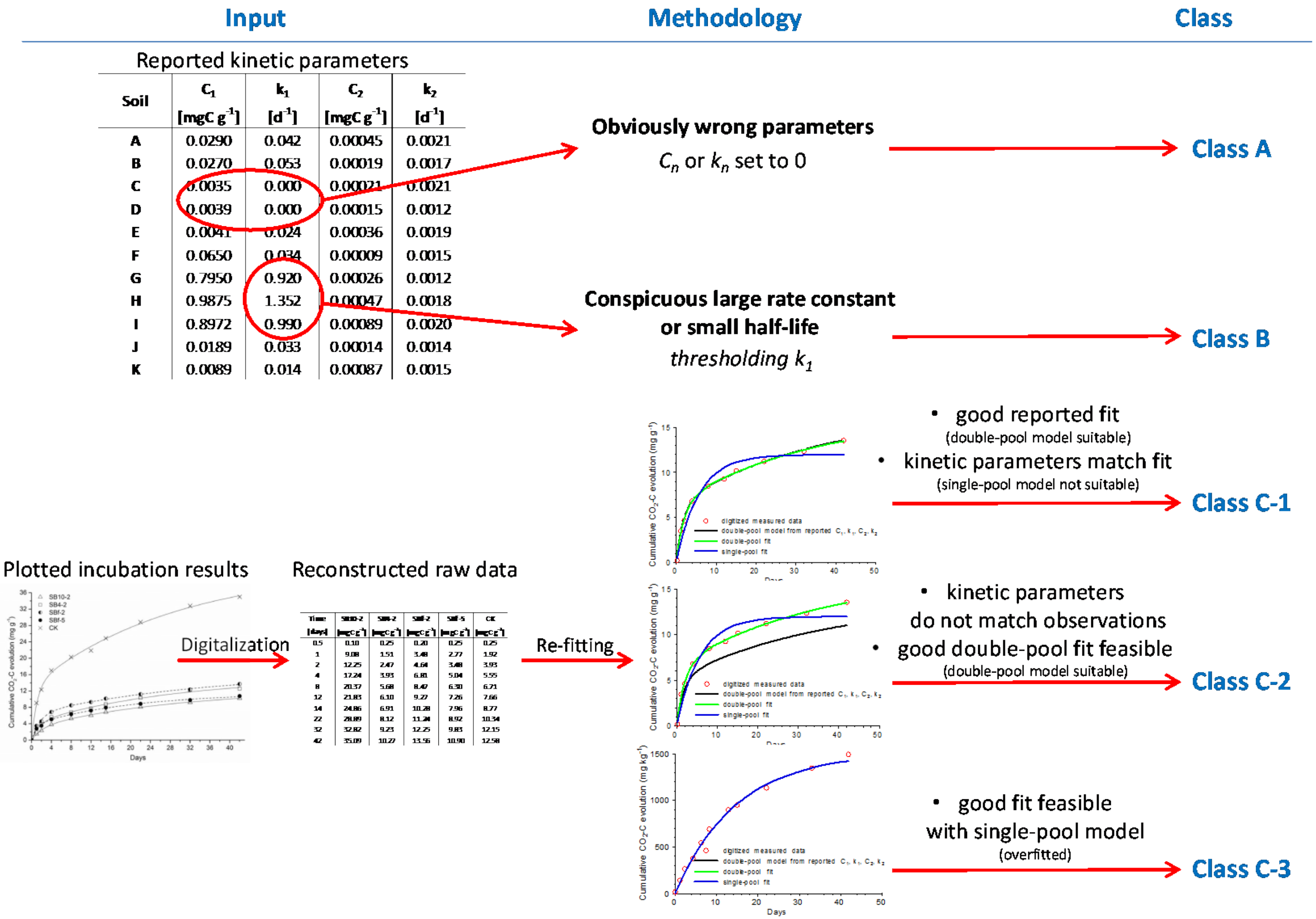

Figure 9: Flowchart for the data base analysis.

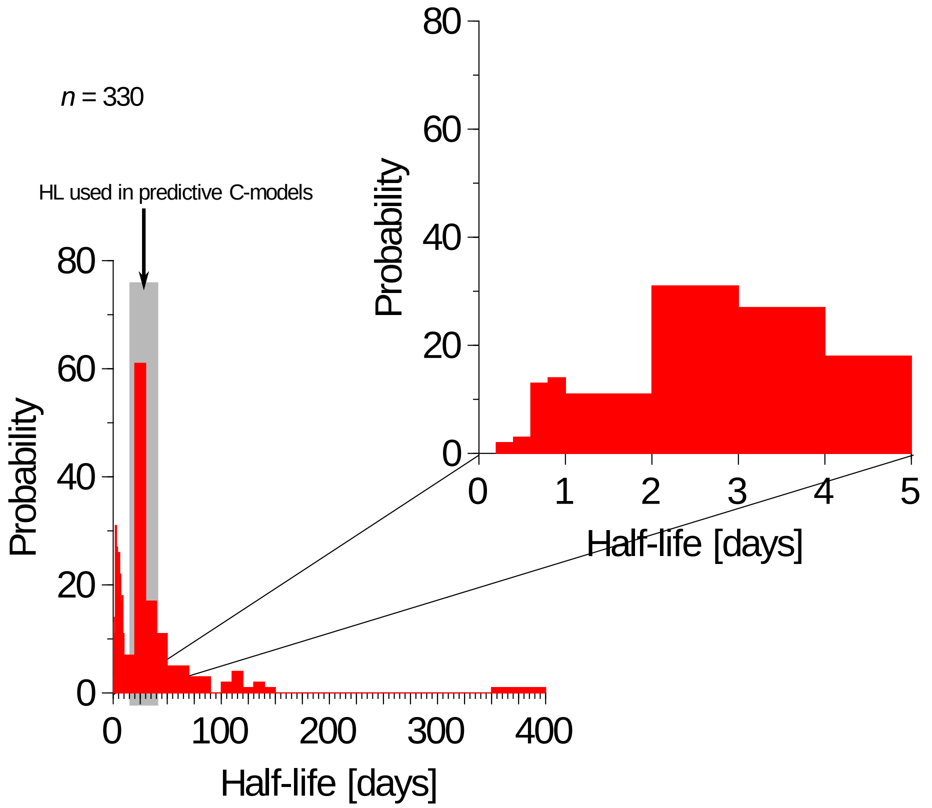

Figure 10: Histogram of half-lives [days] of the fast C-pool extracted from the literature data
975 base. Grey bar indicate span of half-lives used in complex predictive carbon models.

Appendix 1

Additional statistical measures for the goodness of fits

The first step during the fitting process should always be the model identification (single-pool model versus double-pool model). Typically the R^2 is being used for the judgement for the goodness of fit. Unfortunately, the R^2 value is not an adequate metric for the evaluation for the fitting of non-linear functions (Spiess & Neumeyer, 2010). Against a theoretical background R^2 is not valid for non-linear problems. Using the R^2 in linear regression relates the sum of squares of the regression to the total sum of squares based on the assumption that the total sum of squares equals the sum of squares of regression plus the sum of squared residuals. However, for non-linear regressions this is simply not given. The coefficient of model efficiency ME (Nash & Sutcliffe, 1970) is mathematically identical to the coefficient of determination, relating the sum of squared residuals to the total sum of squares:

$$ME = 1 - \frac{SSR}{\sum_{i=1}^{n}\left(x_{obs,i} - \bar{x}_{obs}\right)^2} \qquad [A1]$$

In contrast to R^2 the ME is defined for values ranging between $-\infty$ and 1, and for ME < 0 the observed mean would be a better predictor than the fitted model. However, the R^2 can be used for log-linear transformed single pool approaches.

Alternatively, an information theory criterion could be applied, like the corrected Akaike information criterion (AIC_c) (Hurwich & Tsai, 1989):

$$AIC_c = AIC + \frac{2(p+1)(p+2)}{n-p-2} \qquad [A2]$$

The corrected AIC_c should be applied since the original AIC (Akaike, 1974), defined as

$$AIC = n \cdot \log_n\left(\frac{SSR}{n}\right) + 2(p+1) \qquad [A3]$$

with the SSR as the sum of squared residuals [Eq. 3]. In general, the AIC is prone to over-fitting for problems with a rather small number of observations n. In most incubation studies the *n/p* ratio would be smaller than 40, which points to the use of the AIC_c. Smallest, i.e. more negative, AIC_c values characterize the best model. Notably, the procedures described above will help to identify the best model, i.e. the model providing the smallest error in relation to a minimum number of model parameters. However, it will not necessarily identify the 'true' model.