Ecology, 91(4), 2010, pp. 1225–1236

Single-pool exponential decomposition models: potential pitfalls in their use in ecological studies

E. Carol Adair,[1,2,4] Sarah E. Hobbie,[2] and Russell K. Hobbie[3]

[1]*Department of Forest Resources, University of Minnesota, Saint Paul, Minnesota 55108 USA*
[2]*Department of Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, Minnesota 55108 USA*
[3]*Department of Physics, Emeritus, University of Minnesota, Minneapolis, Minnesota 55455 USA*

Abstract. The importance of litter decomposition to carbon and nutrient cycling has motivated substantial research. Commonly, researchers fit a single-pool negative exponential model to data to estimate a decomposition rate (k). We review recent decomposition research, use data simulations, and analyze real data to show that this practice has several potential pitfalls. Specifically, two common decisions regarding model form (how to model initial mass) and data transformation (log-transformed vs. untransformed data) can lead to erroneous estimates of k. Allowing initial mass to differ from its true, measured value resulted in substantial over- or underestimation of k. Log-transforming data to estimate k using linear regression led to inaccurate estimates unless errors were lognormally distributed, while nonlinear regression of untransformed data accurately estimated k regardless of error structure. Therefore, we recommend fixing initial mass at the measured value and estimating k with nonlinear regression (untransformed data) unless errors are demonstrably lognormal. If data are log-transformed for linear regression, zero values should be treated as missing data; replacing zero values with an arbitrarily small value yielded poor k estimates. These recommendations will lead to more accurate k estimates and allow cross-study comparison of k values, increasing understanding of this important ecosystem process.

Key words: *data transformation; decomposition rate; litterbag; model fitting; regression.*

Introduction

The fundamental importance of litter decomposition to global carbon and nutrient cycles has inspired the development of various mathematical models that describe and predict patterns of litter mass loss. Of these, the single-pool model is most commonly used (Olson 1963):

$$M(t) = M(0)e^{-kt} + \varepsilon \quad (1)$$

where $M(t)$ is mass remaining at time t, $M(0)$ is initial mass, k is the single rate at which all litter constituents are assumed to decompose, and ε is random error assumed to be normal with mean = 0 and standard deviation, SD. This model is often fit to data using the method of least squares (LS), which minimizes Q, the mean square error:

$$Q = \frac{1}{N}\sum_i w_i[M(t_i) - M(t_i)_{\text{prdc}}]^2 \quad (2)$$

where N is the number of data points, predicted mass, M_{prdc}, is calculated using Eq. 1, and weights, w_i, may be set equal to 1 or to estimates of each measurement's precision (Hobbie and Roth 2007). Because this method requires nonlinear LS, investigators often log-transform the data and Eq. 1 in order to use linear LS programs. However, because ε in Eq. 1 is additive, there is no arithmetic process that can transform Eq. 1 into a linear form; researchers must therefore assume that ε is multiplicative to use linear LS (Talpaz et al. 1981, Smith 1993, Packard and Boardman 2009):

$$M(t) = M(0)e^{-kt} \times e^{\varepsilon}. \quad (3)$$

Unlike Eq. 1, Eq. 3 may be log-transformed for linear LS programs as

$$\ln[M(t)] = \ln[M(0)] - kt + \varepsilon \quad (4)$$

with mean square error

$$Q = \frac{1}{N}\sum_i w_i\left\{\ln[M(t_i)] - \ln[M(t_i)_{\text{prdc}}]\right\}^2. \quad (5)$$

Workers in other fields have noted that linearizing similar equations does not provide the same parameter estimates as their nonlinear counterparts (Talpaz et al. 1981, Bailer and Portier 1990, Smith 1993, Packard and Boardman 2009), but in decomposition research this remains an under-recognized consequence of substituting Eq. 4 for Eq. 1 (particularly if all $w_i = 1$). Here, we (1) briefly review the assumptions underlying the use of transformed (linear LS) vs. untransformed (nonlinear LS) data, (2) present a

Manuscript received 13 March 2009; revised 24 June 2009; accepted 2 July 2009. Corresponding Editor: P. M. Groffman.

[4] Corresponding author. Present address: National Center for Ecological Analysis and Synthesis (NCEAS), University of California, 735 State Street, Suite 300, Santa Barbara, California 93101-5504 USA. E-mail: adair@nceas.ucsb.edu

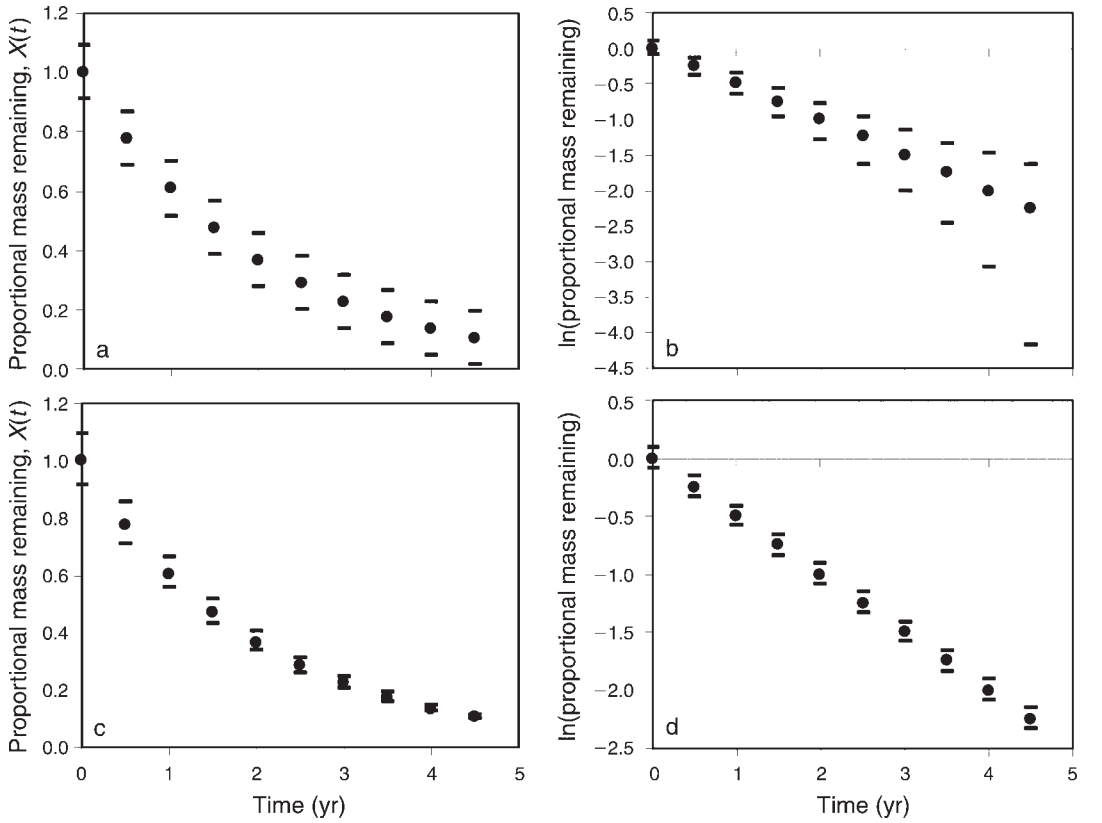

FIG. 1. Simulated data of proportional mass remaining, $X(t)$, with mean decomposition rate, k, = 0.5. Error bars in panels (a) and (b) show constant error [$X(t) \pm 0.09$]. Error in panels (c) and (d) is constant fractional [lognormal; $X(t) \times \exp(\pm 0.09)$]. Data in panels (a) and (c) are untransformed; data in panels (b) and (d) are log-transformed. The decomposition rate, k, is used in Eq. 7 [$X(t) = 1e^{-kt}$], to generate data for the proportional mass remaining (e.g., for leaf litter mass decomposing over time). To these values, constant or constant fractional error was added as shown. Mass declines over time as it is decomposed.

literature review summarizing the manner in which model fitting and data transformation are currently handled in decomposition studies, and (3) use simulated and real data to demonstrate how data transformation and regression choice affect estimates of k.

The repercussions of log-transforming data for linear regression

One may intuitively expect estimates of k from nonlinear LS and Eq. 1 and linear LS and Eq. 4 to be interchangeable, but the estimates are often quite different, because using untransformed data (nonlinear LS) and log-transformed data (linear LS) imply different error structures and weights (w_i; Talpaz et al. 1981, Bailer and Portier 1990, Packard and Boardman 2009). Constant w_i's (all $w_i = 1$) in Eq. 2 imply constant variance across data points and all points have equal influence estimating k (additive error; Fig. 1a). If data with constant variance are log-transformed for linear LS (all $w_i = 1$), variance is nonconstant and estimates of k are influenced more by small than by large values (Fig. 1b). Constant variance for log-transformed data (Fig. 1d) indicates constant fractional variance for the raw data (multiplicative error; Fig. 1c). This is what is implied by log-transforming data and using linear LS (all $w_i = 1$). If data with constant fractional variance are untransformed for nonlinear LS, variance is (less severely) nonconstant and estimates of k are influenced slightly more by large than by small values (Fig. 1c). Thus, choosing an inappropriate data transformation and LS method may produce erroneous estimates of k.

Data with nonconstant variance may also be fit using weighted LS regression, which attempts to give each data point the proper amount of influence over parameter estimates by weighting each measurement according to its precision (NIST 2008). However, estimating w_i's with small numbers of replicates (e.g., <10) can produce poor results (Carroll and Ruppert 1988, NIST 2008). For example, without sufficient replication, the common procedure for estimating w_i's, w_i = 1/replicate variance, yields extremely variable weights and results that do not properly control the influence of measurements on parameter estimates (Carroll and Ruppert 1988, NIST 2008). Decomposition data typically have few replicates, so this method likely yields inappropriate w_i's. We therefore suggest using this method only when replicate numbers are fairly large, so

that w_i's may be estimated precisely (Carroll and Ruppert 1988, NIST 2008).

An additional consequence of fitting log-transformed data using Eq. 4 (linear LS) is that the de-transformed equation will provide biased mass-remaining predictions (de-transformed predictions will also be biased; Jansson 1985, Smith 1993, Packard 2009). In the log scale, Eq. 4 predicts the arithmetic means of the mass remaining at each time point; de-transforming Eq. 4 into Eq. 3 instead predicts the geometric means, which are biased under-predictions of the arithmetic means (Packard 2009, Packard and Boardman 2009). Such de-transformations are not common practice in decomposition studies, but using estimates of k from linearized models in nonlinear models (e.g., in predictive decomposition or ecosystem models) will provide biased mass-remaining values.

The proportional mass remaining at time zero is one by definition

Fitting single-pool models also requires deciding how to model $M(0)$. Commonly, $M(t)$ is normalized by $M(0)$ to calculate the proportional mass remaining, $X(t)$:

$$X(t) = \frac{M(t)}{M(0)}. \quad (6)$$

In this case, proportional initial mass, $X(0) = M(0)/M(0)$. Thus, $X(0)$ equals 1 by definition, and Eq. 1, written in terms of the proportional mass remaining, becomes

$$X(t) = 1e^{-kt}. \quad (7)$$

Similarly, Eq. 4 becomes

$$\ln[X(t)] = \ln[1] - kt = 0 - kt. \quad (8)$$

Because $X(0) = 1$ by definition, neither $X(0)$ nor $\ln[X(0)]$ should be estimated as model parameters. Instead, $X(0)$ should be set equal to one in the model equation (as in Eqs. 7 and 8). Many linear regression procedures in statistical programs automatically estimate a y-intercept unless otherwise instructed, so caution must be used to prevent such procedures from implicitly allowing an $X(0)$ that is greater or less than one and therefore biologically unreasonable.

Methods

Decomposition literature review

To determine how researchers currently deal with data transformation and model-fitting issues, we searched the ISI Web of Science for articles published from January 2002 to February 2008 that had the terms litter, decomposition, and rate(s) in the abstract, key words, or key words plus. Of the 1538 articles found by this search, 498 were fine-root or leaf litterbag studies available from the University of Minnesota's electronic journal collection. For the studies that fit a single-pool model to their data (the majority; see *Results*), we used the following rules to determine which method, log-transformed (LT) with linear LS or untransformed (UT) with nonlinear LS, was used:

1) If, despite presentation of a nonlinear (N) equation in the methods, results indicated that the linear (L) model was used on LT data, we counted the model as "L-LT." When the linear model was given in the methods, we never found that nonlinear regression on untransformed data was used instead. Thus, such studies were recorded as "L-LT."

2) If a nonlinear model was given in the methods, but it was not made clear that nonlinear LS was used on the UT data, we counted it as an "assumed N-UT."

3) Studies that explicitly stated nonlinear regression was used or reported parameter values within a nonlinear equation in the figures, tables, or results were recorded as "N-UT."

4) If no information about the model or data transformation used was given (e.g., "breakdown rates were computed using an exponential decay model"), we recorded it as "unknown."

We used the following rules to categorize how the initial value, $M(0)$, was modeled:

1) Studies were recorded as "$M(0)$ unknown" if it was not explicitly stated whether $M(0)$ was estimated and either (a) the equation given in the methods included a parameter for $M(0)$ or (b) no equation was given in the methods. Because linear LS programs often estimate an intercept by default, using this rule likely underestimates the "$M(0)$ estimated" (rule 2, below).

2) If the given equation did not have an explicit parameter for $M(0)$, we assumed that the value was fixed at one (e.g., if $X(t) = e^{-kt}$, then $X(0) = 1$). These cases were recorded as "assumed fixed $M(0)$." However, "assumed fixed $M(0)$" cases were changed to "$M(0)$ estimated" when figures or tables presented in the results clearly indicated that $M(0)$ was estimated.

3) If authors stated that initial mass was fixed at a certain value, we counted it as "fixed $M(0)$."

4) If authors stated that they estimated $M(0)$, it was recorded as "$M(0)$ estimated."

In addition to the above information, we recorded (1) study length, (2) what decomposition model(s) were fit, and (3) how models were compared or selected if more than one model was fit. The 498 reviewed litterbag studies are listed in Appendix A.

Data simulations for single-pool k *estimation*

We used four data simulation runs to investigate the consequences of estimating single-pool k using transformed data (linear LS) and untransformed data (nonlinear LS) methods. Each run increased the length of the simulation from 200, to 400, to 600, to 1500 days using the following time series: 30, 60, 100, 150, 200, 250, 300, 350, 400, 500, 600, 700, 1000, 1500 days (4.1 years). Increasing the length of a run concurrently decreased the mass remaining at the end of the run (Fig. 2). Log-transforming normal data gives small values more influence on parameter estimates than large values, so

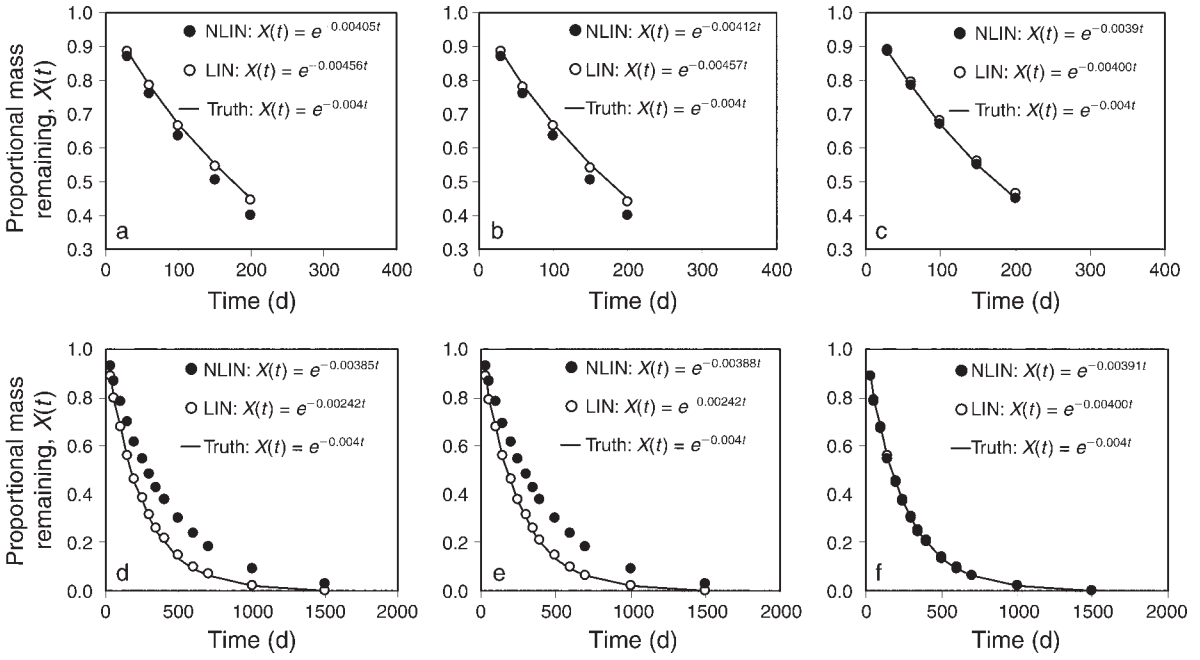

FIG. 2. Proportional mass remaining, $X(t)$, predicted by mean decomposition rate, k, estimates from nonlinear (NLIN; untransformed data) and linear regression (LIN; log-transformed data) from (a–c) 200-day and (d–f) 1500-day simulations. Errors are normal (a and d), truncated normal (b and e), or lognormal (c and f). Proportional initial mass, $X(0) = 1$ for all simulations. "Truth" is the proportional mass remaining, calculated using Eq. 7 with $k = 0.004$ d^{-1}, prior to the addition of normal, lognormal, and/or truncated error.

varying time series length allowed us to investigate how small mass-remaining values affected the ability of each transformation and regression technique to accurately estimate k.

In each simulation run we used the true proportional mass remaining at each time point, calculated using Eq. 7 and $k = 0.004$ d^{-1}, to create three data sets with different error structures: normal, lognormal, and truncated normal. We expected that k would be accurately estimated by unweighted nonlinear LS of untransformed data if the data had normal, constant errors or by unweighted linear LS of log-transformed data if data had lognormal, constant fractional error. Normal errors were generated using a normal distribution (mean = 0, SD = 0.2, a typical SD for mass-remaining data; Harmon 2007, Hobbie 2008) and added to each true mass-remaining value. Because negative mass-remaining values are biologically unrealistic, generated negative values were set = 0. Within a given time series, only the negative values were set = 0 (i.e., the occurrence of a zero value did not force all remaining values to be set = 0). To estimate k with linear LS, zero values were set = 0.00001 and data were log-transformed. We compared results from these simulations with those from runs treating zero values as missing data. We generated lognormal error by multiplying true mass-remaining values by exp(normal error).

While errors are often assumed to be lognormal or normal, we found that errors in two decomposition data sets were instead small for high (near 1) and low (near 0) proportional mass-remaining values and large for mid-range mass-remaining values (Appendix B; Harmon 2007, Hobbie 2008). To determine how well each data transformation and LS method estimated k for such data, we generated data with "truncated normal errors" by adding normal error to true mass-remaining values, setting negative values = 0 and setting values greater than 1.05 = 1. This generated errors similar to those found in the real data sets we examined (Appendix B).

Each simulation run estimated k for 10 000 normal, truncated normal, and lognormal data sets using nonlinear LS (untransformed data) and linear LS (log-transformed data). We estimated k for the mean of three replicates per time point. We reran the simulations using one replicate, but results were nearly identical and are not presented. To determine how well each regression and data transformation method estimated k, we calculated the mean, standard error (SE), bias,

$$\%\text{Bias} = \sum_i [(\text{estimated } k_i - \text{true } k)/\text{true } k]/10\,000 \times 100 \tag{9}$$

and relative error (RE),

$$\%\text{RE} = \sum_i [\text{abs}(\text{estimated } k_i - \text{true } k)/\text{true } k]/10\,000 \times 100 \tag{10}$$

for the k estimates from each simulation, where $i = 1$ to

TABLE 1. Results of the literature search: models used in litterbag studies from January 2002 to February 2008.

Model type	Total	Percentage of total
Linear	22	4.42
Single	319	64.06
Asymptotic	5	1.00
Double	20	4.02
Other	21	4.22
Model comparisons (explicit)	35 (26)	7.01 (5.21)
No model fit	155	31.12
Total	498	

Notes: Percentage values are percentage of total litterbag studies (out of 498 studies). "Explicit" model comparisons described and statistically compared all the models that were fit to the data.

10 000 k estimates per fitting method per simulation (abs, absolute value). Bias provided a measure of each method's over- or underestimation of true k, while RE provided a measure of the magnitude of the difference between true k and each k estimate.

To investigate the consequences of estimating proportional initial mass, $X(0)$, vs. fixing it at the true value, $X(0) = 1$, we repeated the simulations, estimating k and $X(0)$ for each data set using both data transformation and regression methods. For log-transformation, zero values were treated as missing data. We calculated the mean, SE, bias, and RE for all k and $X(0)$ estimates.

Finally, because work with similar equations suggests that (1) the amount of error in the data strongly influences the ability of nonlinear regression to accurately estimate parameters (Bottcher 2004) and (2) the difference between linear and nonlinear parameter estimates will increase with increasing error (Jansson 1985, Packard and Boardman 2009), we repeated the simulation runs with less (SD = 0.004) and more (SD = 0.33) error than in the original simulations. All simulations and statistical analyses were performed in R version 2.8.1 (R Development Core Team 2008).

Real data analysis

Our real data contained data from Hobbie (2008) plus S. E. Hobbie's unpublished filter paper mass loss data from the same experiment (hereafter, the Hobbie [2008] data set). The Hobbie (2008) experiment was established at Cedar Creek Ecosystem Science Reserve in central Minnesota, USA (45°24′0″ N, 93.2°12′0″ W). Eight litters were decomposed for five years at eight sites (two old fields, a hardwood forest, two oak stands, two pine stands, and an aspen stand), with a nitrogen addition treatment at each site ($n = 6$). Details are presented in Hobbie (2005, 2008). Zero values were treated as missing data for log transformation (see *Results*). Using nonlinear (untransformed data) and linear (log-transformed data) LS, we estimated k with $X(0) = 1$ and while estimating $X(0)$.

We also investigated the ability of weighted nonlinear LS (untransformed data) to accurately estimate k. In the simulations, we used $w_i = 1/\text{variance}$ of the three replicates (Eq. 2), estimating k using (1) the mean mass remaining of the three replicates per time point and (2) only one replicate. In the Hobbie (2008) data we used 1/variance of all six replicates, but estimated k separately for each replicate. This method performed well in the simulations for which variance was relatively constant, but performed poorly in the real data, often producing extremely variable w_i's, unrealistic k values, and predictions that fit the data very poorly (data not shown). Examination of the data and w_i's revealed that the data had too few replicates to calculate accurate w_i estimates. We therefore do not present these results.

As a resource for decomposition researchers, we have provided a simplified version of the R code we used to estimate k using untransformed data and nonlinear regression (see Supplement).

RESULTS

Literature review

Of 498 litterbag studies, 343 fit a model to their data (Table 1). The single-pool model was most commonly used, but linear, double-pool, and asymptotic or limit value models were also used (Table 1; Weider and Lang 1982, Berg and Ekbohm 1991). Infrequently used were Ågren and Bosatta's (1996) Q model, quadratic models (Wieder and Lang 1982), Godshalk and Wetzel's (1978) single-pool decaying decomposition rate model, and models predicting mass loss as a function of litter chemistry and/or climate (e.g., Liski et al. 2003). Few studies fit more than one model to data and often no information was given about what models were com-

TABLE 2. Results of the literature search: how the single-pool negative exponential model is used to estimate litter decomposition rates (k) and/or initial mass ($M(0)$).

Method	No. studies	Percentage of total
Litterbag studies		
Total	498	
Calculated single-pool k	319	
Initial mass, $M(0)$		
$M(0)$ unknown	183	57.37
Fixed $M(0)$ at 100% or initial mass	30	9.40
Assumed fixed $M(0)$	70	21.94
$M(0)$ estimated using data†	37 (26)	11.59 (70.03)
Data transformation and regression		
Unknown	61	19.12
N-UT	48	15.05
Assumed N-UT	87	27.27
L-LT	126	39.50

Notes: Percentage values are percentages of total litterbag studies that estimated k using the single-pool model (319 studies). Abbreviations are: N-UT, authors explicitly stated that they used nonlinear regression on untransformed data; Assumed N-UT, studies in which it was not explicitly stated, but only implied that the authors used nonlinear, untransformed data (see *Methods*); L-LT, linear regression on log-transformed data.

† $M(0)$ range reported.

TABLE 3. Results from simulation runs investigating the ability of each data transformation and regression technique combination to accurately estimate the single-pool decomposition rate, k, with proportional initial mass, $X(0) = 1$, and non-positive mass-remaining values treated as missing data (log-transformed normal and truncated normal data only).

Measurement, by simulation period	Normal errors		Truncated normal errors		Lognormal errors	
	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
200 days						
Bias (%)	1	14	3	14	−4	0
Relative error (%)	15	22	15	22	10	8
Mean k	0.00405	0.00456	0.00412	0.00457	0.00385	0.00400
0.00001 mean k	NA	0.00490	NA	0.00492	NA	NA
Mean MR	0.450	0.402	0.452	0.404	0.459	0.450
400 days						
Bias (%)	−1	8	0	8	−3	0
Relative error (%)	10	15	10	15	5	3
Mean k	0.00397	0.00433	0.00400	0.00432	0.00390	0.00400
0.00001 mean k	NA	0.00631	NA	0.00629	NA	NA
Mean MR	0.218	0.193	0.218	0.195	0.206	0.202
600 days						
Bias (%)	−3	−3	−2	−3	−2	0
Relative error (%)	9	13	9	13	4	2
Mean k	0.00390	0.00386	0.00394	0.00386	0.00391	0.00400
0.00001 mean k	NA	0.00684	NA	0.00686	NA	NA
Mean MR	0.133	0.138	0.133	0.138	0.093	0.091
1500 days						
Bias (%)	−4	−40	−3	−40	−2	0
Relative error (%)	9	40	9	40	4	1
Mean k	0.00385	0.00242	0.00388	0.00242	0.00391	0.00400
0.00001 mean k	NA	0.00514	NA	0.00516	NA	NA
Mean MR	0.082	0.107	0.081	0.106	0.003	0.002

Notes: Estimated k values are from simulations using three replicates per time point (true $k = 0.004$ d^{-1}). Each simulation run generated 10 000 data sets with normal (constant), truncated normal (constant), and lognormal (constant fractional) errors, which were used to estimate k using linear regression of log-transformed data and nonlinear regression of untransformed data. Mean k standard errors were all ≤ 0.00001. Mean MR is the proportional mass remaining at the end of the simulation; "0.00001 mean k" is the mean k from simulations in which non-positive mass-remaining values were replaced by 0.00001. "NA" indicates not applicable.

pared or how (Table 1). Detailed model comparisons used R^2, statistical significance, or mean square error as model selection criteria. When models were not fit, data were usually analyzed with repeated-measures ANOVA or ANOVAs or t tests on data at one to many time point(s).

When fitting the single-pool model, all methods and model forms were used, but the amount of method detail provided varied widely. The largest fraction of studies estimated k using log-transformed data and linear LS (40%; Table 2). Only 15% of studies explicitly stated that they used nonlinear LS on untransformed data to estimate k (Table 2). For the remaining 45%, we either had to assume a method or it was unknown. Most often, it was also unclear whether or not initial mass was estimated (57%; Table 2). Often, we assumed that initial mass was fixed (based on the equation provided in the methods), although no explicit information about initial mass estimation was given (22%; Table 2). In rare cases, it was explicitly stated that initial mass was fixed (9%) or

TABLE 4. Results from simulation runs investigating the effect of choosing an arbitrary replacement value for zero values on estimating the single-pool decomposition rate, k, using linear regression on log-transformed data with normal and truncated normal errors (true $k = 0.004$ d^{-1}; $X(0) = 1$).

Zero replacement value	Normal errors				Truncated normal errors			
	Bias (%)	Relative error (%)	Decomposition, k		Bias (%)	Relative error (%)	Decomposition, k	
			Mean	SE			Mean	SE
0.01	−17	17	0.00331	3.73×10^{-6}	−17	17	0.00332	3.71×10^{-6}
0.00001	28	31	0.00514	9.04×10^{-6}	29	31	0.00516	9.10×10^{-6}
0.0000001	97	97	0.00788	1.75×10^{-5}	97	97	0.00787	1.75×10^{-5}

Notes: All simulation runs were 1500 d. Mean k and SE are for the 10 000 k estimates within one run.

TABLE 5. Results from simulation runs investigating the ability of each transformation and regression technique to accurately estimate the single-pool decomposition rate, k, when proportional initial mass, $X(0)$, was also estimated as a model parameter (i.e., $X(0)$ was not fixed at 1).

Measurement, by simulation period	Normal errors		Truncated normal errors		Lognormal errors	
	Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
200 days						
Bias (k) (%)	1	13	−5	8	0	0
Relative error (k) (%)	28	37	26	36	18	17
Mean k	0.00404	0.00451	0.00378	0.00434	0.00401	0.00399
Mean $X(0)$	1.01	1.01	0.97	0.98	1.02	1.00
400 days						
Bias (k) (%)	−2	3	−5	2	0	0
Relative error (k) (%)	15	22	15	22	9	6
Mean k	0.00394	0.00414	0.00381	0.00409	0.00401	0.00400
Mean $X(0)$	1.00	0.97	0.97	0.95	1.02	1.00
600 days						
Bias (k) (%)	−4	−15	−7	−15	0	0
Relative error (k) (%)	14	24	14	24	7	4
Mean k	0.00383	0.00341	0.00372	0.00340	0.00402	0.00400
Mean $X(0)$	0.99	0.86	0.96	0.85	1.02	1.00
1500 days						
Bias (k) (%)	−7	−60	−10	−60	0	0
Relative error (k) (%)	15	60	15	60	7	2
Mean k	0.00372	0.00161	0.00360	0.00159	0.00401	0.00400
Mean $X(0)$	0.98	0.55	0.95	0.54	1.02	1.00

Notes: Each simulation run generated 10 000 data sets with normal, truncated normal, and lognormal errors. These data sets were used to estimate k and $X(0)$ using linear regression on the log-transformed data and nonlinear regression on the untransformed data. The mean of three replicates per time point (true $k = 0.004\ d^{-1}$) was used to estimate k and $X(0)$. Mean k SEs were all ≤ 0.00002. Mean $X(0)$ SEs were all ≤ 0.002.

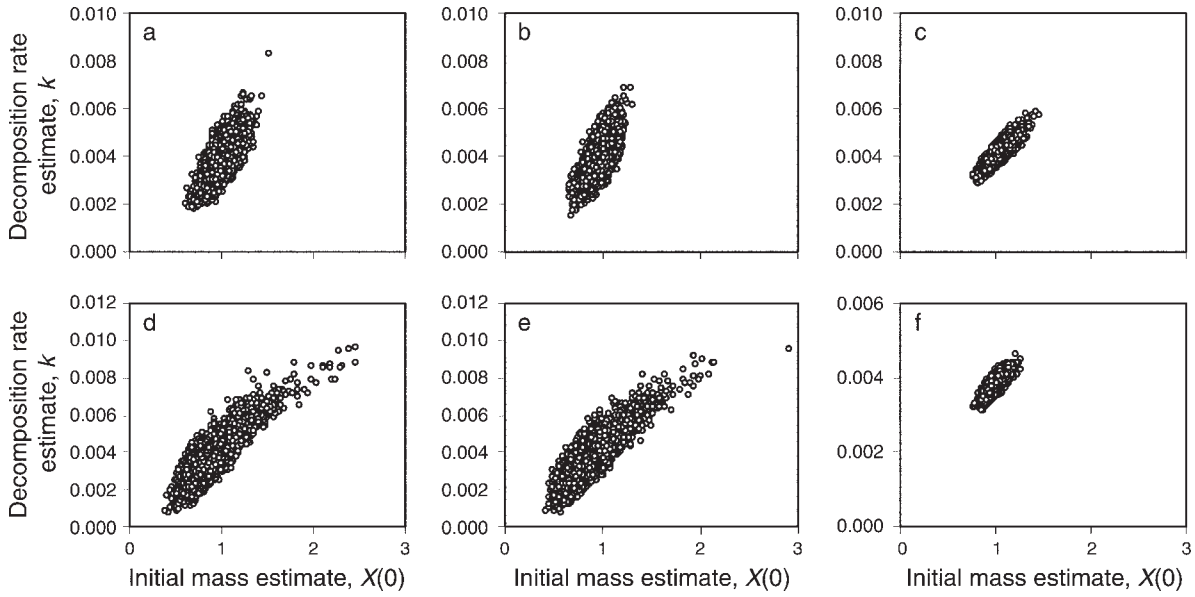

FIG. 3. Decomposition rate, k, vs. proportional initial mass estimates, $X(0)$, from nonlinear regression of untransformed data with (a) normal, (b) truncated normal, and (c) lognormal errors, and from linear regression of log-transformed data with (d) normal, (e) truncated normal, and (f) lognormal errors. Results are from the 600-day simulation but were similar for all simulations.

TABLE 6. Results from simulation runs investigating the effect of changing the standard deviation (SD) of simulated data on single-pool decomposition rate estimates, k, with fixed and estimated proportional initial mass, $X(0)$.

Simulation period (d)	SD	Bias or RE	$X(0)$ fixed at 1					
			Normal errors		Lognormal errors		Truncated normal errors	
			Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
200	0.33	Bias (%)	0	21	−10	0	10	24
		RE (%)	24	35	17	14	25	36
	0.2	Bias (%)	1	14	−4	0	3	14
		RE (%)	15	22	10	8	15	22
	0.07	Bias (%)	0	2	0	0	0	1
		RE (%)	5	6	3	3	5	6
400	0.33	Bias (%)	−5	−2	−7	0	−1	−1
		RE (%)	15	18	9	5	15	18
	0.2	Bias (%)	−1	8	−3	0	0	8
		RE (%)	10	15	5	3	10	15
	0.07	Bias (%)	0	3	0	0	0	3
		RE (%)	3	5	2	1	4	6
600	0.33	Bias (%)	−8	−16	−6	0	−5	−17
		RE (%)	15	21	8	4	14	21
	0.2	Bias (%)	−3	−3	−2	0	−2	−3
		RE (%)	9	13	4	2	9	13
	0.07	Bias (%)	0	5	0	0	0	4
		RE (%)	3	8	1	1	3	7
1500	0.33	Bias (%)	−11	−51	−6	0	−8	−51
		RE (%)	17	51	8	2	15	51
	0.2	Bias (%)	−4	−40	−2	0	−3	−40
		RE (%)	9	40	4	1	9	40
	0.07	Bias (%)	0	−22	0	0	0	−22
		RE (%)	3	23	1	0	3	23

Notes: The SD was either smaller (0.067) or larger (0.33) than in the initial simulations (0.2). Bias and relative error (RE) for the estimates of k are shown. Each simulation run generated 10 000 data sets with normal, truncated normal, and lognormal errors, which were used to estimate k or k and $X(0)$ using linear regression on the log-transformed data and nonlinear regression on the untransformed data. The mean of three replicates per time point (true $k = 0.004$ d^{-1}) was used to estimate the model parameters.

estimated (12%; Table 2). In the latter case, 70% of studies reported the estimates.

Data simulations

Nonlinear regression of untransformed data accurately estimated k for data with all error types while linear regression of log-transformed data only consistently estimated k accurately for data with lognormal errors (Fig. 2, Table 3). Replacing zero mass-remaining values with an arbitrarily small replacement value (RV) to allow log transformation and linear regression of normal data (a common practice) greatly influenced k estimates (Tables 3 and 4). Using an RV that is very different from measured values (e.g., 0.00001 vs. 0.002 at $t = 1500$) likely creates influential outliers (Cook and Weisberg 1999). The influence of these points is further amplified because log-transforming normal data causes small values to have more influence than large values (Fig. 1b; this was not a problem for generated lognormal data, which only contained values > 0). Thus, if the RV was small compared to mass-remaining values at the same time point (e.g., 0.0000001 vs. 0.002), disproportionate weighting of the RV resulted in an overestimate of k (Table 4; Appendix C). If the RV was large compared to mass-remaining values at the same time point (e.g., 0.01 vs. 0.002), k was underestimated (Table 4; Appendix C). Results were nearly identical for data with truncated errors (Table 4). Treating zero values as missing data resulted in more accurate linear k estimates (Table 3). Thus, we report results from simulations treating zero values as missing data.

For data with normal and truncated normal errors, estimating $X(0)$ and k as model parameters yielded less accurate estimates of k than did fixing $X(0) = 1$ in the equation (i.e., using Eqs. 7 or 8) and estimating k: across fitting methods, relative error (RE) and bias were either similar or greater when $X(0)$ was estimated than when it was fixed (Tables 3 and 5). Mean $X(0)$ for data with normal and truncated normal errors ranged from 0.54 to 1.01 (Table 5). In contrast, mean $X(0)$ for data with lognormal errors ranged from 1.00 to 1.02 (Table 5). However, across all simulations and fitting methods, the mean range of k values was nearly two times greater when $X(0)$ and k were estimated (0.0077 d^{-1}) than when only k was estimated (0.0043 d^{-1}), and estimates of k increased with $X(0)$ estimates (Fig. 3). Because estimating $X(0)$ did not yield accurate estimates of k, we focus on simulations that fixed $X(0) = 1$ and only estimated k.

With $X(0) = 1$, nonlinear regression of untransformed data with normal and truncated normal errors accurately estimated k in all simulations, at worst underestimating k by 4% (Fig. 2, Table 3). Log-transforming

TABLE 6. Extended.

$X(0)$ estimated					
Normal errors		Lognormal errors		Truncated normal errors	
Nonlinear	Linear	Nonlinear	Linear	Nonlinear	Linear
0	5	1	−1	−14	−5
46	59	31	28	42	57
1	13	0	0	−5	8
28	37	18	17	26	36
0	1	0	0	0	1
10	11	6	6	10	11
−8	−19	1	0	−17	−23
24	34	15	11	25	36
−2	3	0	0	−5	2
15	22	9	6	15	22
0	4	0	0	0	4
5	9	3	2	5	9
−14	−37	1	0	−21	−40
24	41	12	7	26	43
−4	−15	0	0	−7	−15
14	24	7	4	14	24
0	6	0	0	0	6
5	12	3	1	5	12
−19	−72	1	0	−27	−73
27	73	12	3	30	73
−7	−60	0	0	−10	−60
15	60	7	2	15	60
−1	−36	0	0	−1	−36
5	37	2	1	5	37

data with normal or truncated normal errors for linear LS yielded the least accurate estimates of k, overestimating k by 14% in short simulations and underestimating k by 3–40% in long simulations, due to the large influence of small values (Fig. 2, Table 3). In short simulations, overestimates of k and the corresponding underestimates of mass remaining were likely the result of (1) small values at the end of the time series increasing estimates of k and (2) the underestimation bias introduced by de-transforming the equation that was fit to log-transformed data (Jansson 1985). As simulation length increased and mass remaining decreased, more zero values were generated and treated as missing, leaving only relatively large small values (compared to the true mass remaining). The large influence of these values (due to log-transforming data with normal errors) forced linear regression to increasingly underestimate k as simulation length increased, an effect that was amplified by increasing the data SD (Table 6) and that increased with time series length for a given true k (Appendix D). Linear regression of log-transformed data with lognormal errors and nonlinear regression of the same untransformed data produced accurate estimates of k (Fig. 2, Table 3). Thus, data transformation and regression method choice was much less important for data with lognormal (constant fractional) errors.

Regardless of whether or not $X(0)$ was estimated, increasing the SD of simulated data always increased RE and either had no effect on or increased the bias of estimates of k (Table 6). Similarly, decreasing the SD decreased RE and either had little effect on or decreased the bias of estimates of k (Table 6). For data with normal and truncated normal errors, increasing the SD had a smaller effect on estimates of k if nonlinear regression (untransformed data), rather than linear regression (log-transformed data), was used (Table 6). The opposite was true for data with lognormal errors, but RE and bias remained relatively low (Table 6). As expected, increasing the SD increased the difference between linear and nonlinear estimates of k (Appendix E).

Real data

Estimating $X(0)$ and k for the Hobbie (2008) data resulted in dramatic over- and underestimates of $X(0)$ (Fig. 4). For nonlinear regression of untransformed data, $X(0)$ ranged between 0.49 and 1.8, and for linear regression of log-transformed data, $X(0)$ ranged between 0.35 and 4.06. Given these misestimates and our simulation results, we focus on the results of the fixed $X(0)$ analyses.

Using untransformed (nonlinear LS) and log-transformed (linear LS) data resulted in very different k estimates (Fig. 5). Linear and nonlinear k estimates corresponded well for slowly decomposing litters with low k's, but for rapidly decomposing litters, linear k's were much higher than nonlinear k's (Fig. 5). In general, for substrates with the smallest amount of mass remaining at the end of the experiment (less than approximately 10%), linear k values were substantially larger than nonlinear k values (Fig. 5b). For example, on average, filter paper had the least amount of mass remaining at the end of the experiment and, on average, the linear k estimate was more than 1.5 times greater than the nonlinear k estimate (Appendix F). These results are consistent with our data simulations, which found linear LS to overestimate k when data with constant normal errors had small values at the end of the time series. Conversely, linear LS estimated slightly smaller k's than nonlinear LS for the three next fastest decomposing species, which had nearly twice the amount of mass remaining at the end of the experiment as the fastest decomposing substrate (Appendix F). Our simulation runs generated such results only in long simulations where zero values were treated as missing data, so that the remaining, larger-than-average values had more influence than large values early in the time series (e.g., Fig. 2d). Similarly, in the Hobbie (2008) data, linear k's were smaller than nonlinear k's when the last one or two values were greater than earlier values.

DISCUSSION AND CONCLUSIONS

Our results suggest several ways to avoid the substantial potential pitfalls that surround fitting single-pool k values. First, we assert that it is biologically appropriate to define initial mass as the true,

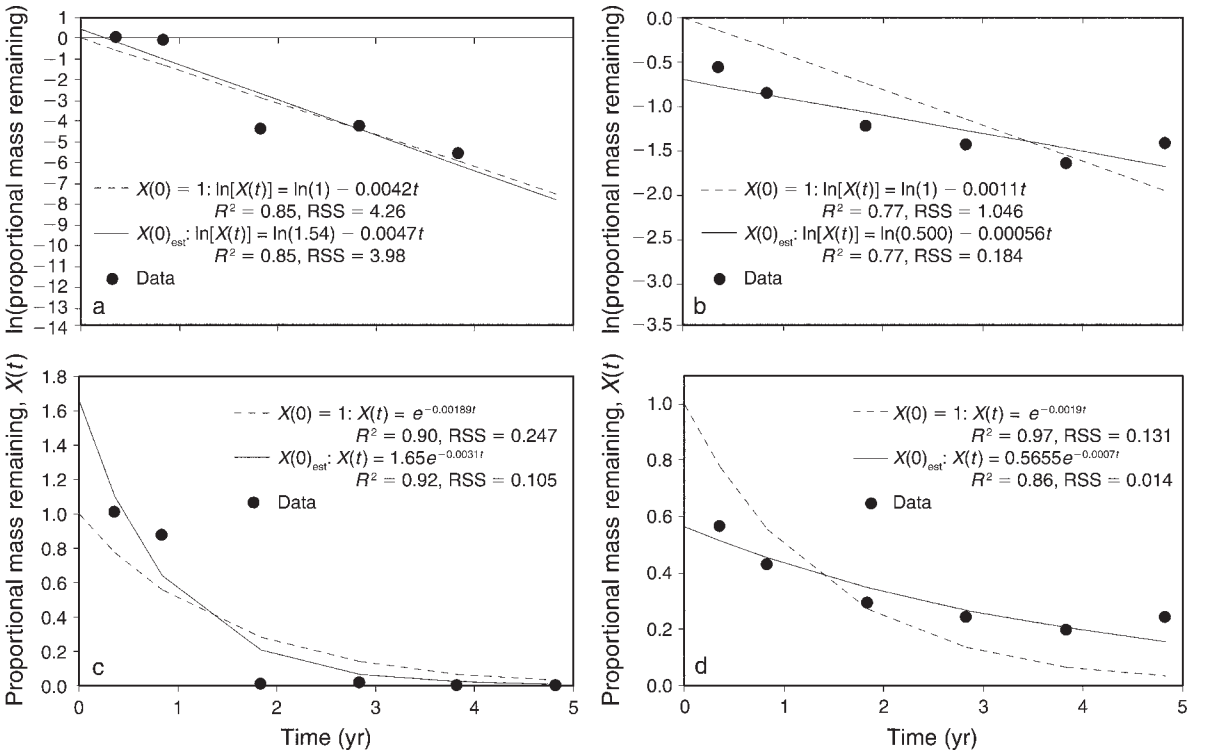

FIG. 4. Single-pool models fit to data of proportional mass remaining, $X(t)$, with proportional initial mass, $X(0)$, estimated or fixed at 1. When $X(0)$ is estimated, $X(0)$ and decomposition rate, k, can be greatly over- or underestimated by (a, b) linear regression of log-transformed data and (c, d) nonlinear regression of untransformed data. Data are two litter types from Hobbie (2008). RSS is the residual sum of squares.

measured initial mass. While estimating initial mass [expressed either as proportional initial mass, $X(0)$, or as the measured, non-normalized initial mass, $M(0)$] and k as model parameters may produce apparently better fits to the data, it produces incorrect initial mass and k values (Fig. 4). In our simulations, over- or underestimates of $X(0)$ led to unrealistically high or low k values (Fig. 3) that are not comparable to k values from studies that do not estimate $X(0)$. Similarly, although we did not address it directly in our simulations, estimating initial mass in the original, measured units, $M(0)$, would lead to the same problems. Therefore, we suggest that if mass-remaining data are expressed proportionally as $X(t)$, $X(0)$ should be set equal to 1 in the model equation (e.g., Eq. 7). If data are modeled in the original units as $M(t)$, $M(0)$ in the model equation should equal

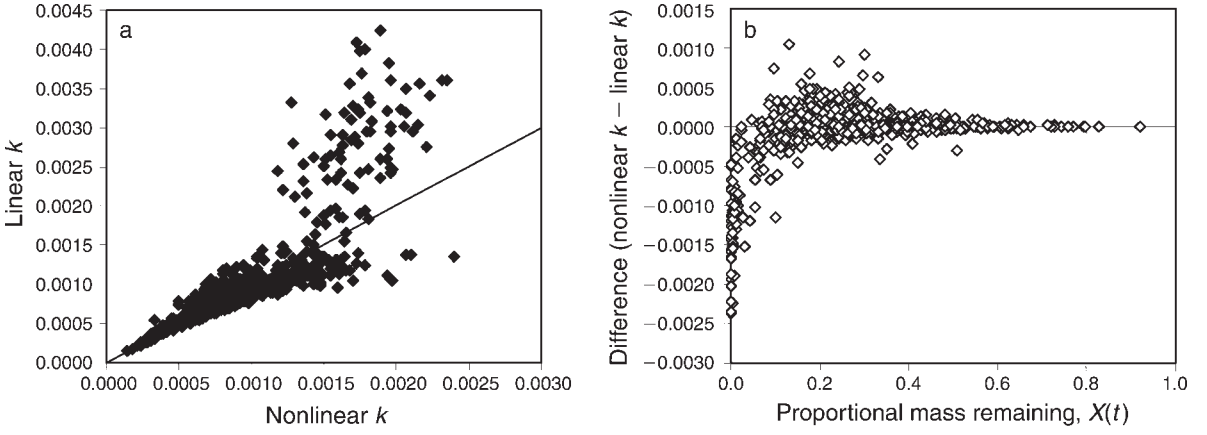

FIG. 5. (a) Linear regression (log-transformed data) of decomposition rate, k, vs. nonlinear (untransformed data) k estimates from the Hobbie (2008) data set and (b) the difference between k estimates from nonlinear and linear regression (nonlinear k − linear k) vs. the proportional mass remaining at the end of the experiment (five years). The line in panel (a) is a 1:1 line.

measured initial masses. Alternatively, the relatively small amount of measurement error associated with initial masses may be accounted for by constraining initial mass estimates to biologically realistic values. Poor single-pool model fits may be revealed by drastic over- or underestimates of initial mass, low R^2 values if initial mass is fixed, or visual examination and may indicate that other models would more accurately describe mass loss (Harmon et al. 2009). In our real data, when estimated $X(0)$ is $<80\%$ of $X(0)$, 57% of these cases were better fit by an asymptotic or double-pool model; 31% were fit equally well by the single-pool, double-pool, and/or asymptotic model (Appendix G). In contrast, these models did not better fit data where $X(0)$ was overestimated.

Second, in agreement with work on the linearization of similar equations via log transformation (e.g., power law equations; Smith et al. 1980, Talpaz et al. 1981, Benedetti and Sebastiani 1996, Packard and Boardman 2009), we assert that nonlinear regression of untransformed decomposition data will nearly always result in more accurate k estimates than linear regression of log-transformed data, because of issues related to data error structure, bias in the de-transformed log scale predictions, and problems associated with the treatment of zero values when log-transforming data for linear regression. When data were untransformed, nonlinear regression accurately estimated k for data with all error structures. When data were log-transformed, linear regression estimated k accurately when errors were lognormal (a constant fraction of mass remaining), but when errors were normal (constant) or truncated normal, linear regression often resulted in sizable over- or underestimates of k that increased substantially with the amount of error introduced into the simulated data. Thus, we suggest that log-transforming data to estimate k with linear regression, the most common choice of authors in our literature search, should be reserved for situations in which the data are known to have constant fractional error; but note that most studies have insufficient replication to determine data error structure. If linear regression is used, zero values should be treated as missing data rather than as arbitrarily small values, as our simulations found that this greatly improved the accuracy of linear k estimates. Note that using nonlinear regression on untransformed data eliminates the issue of how to treat zero values, as zero values can be included in nonlinear k estimates. Using nonlinear regression also avoids the problem of how to deal with de-transformation bias from equations (or predictions) that were fit in the log scale. We suggest that careful selection of fitting methods, as we have described above, will lead to more accurate and comparable k estimates, thereby increasing our understanding of this important ecosystem process.

Acknowledgments

This paper is based upon work supported by the National Science Foundation under the Cedar Creek Long Term Ecological Research program (DEB-0620652, DEB-0716587, and DEB-0322057). We thank Kevin Little and anonymous reviewers for helpful comments on the manuscript, and Mark Harmon for use of the LIDET data set.

Literature Cited

Ågren, G. I., and E. Bosatta. 1996. Theoretical ecosystem ecology: understanding element cycles. Cambridge University Press, Cambridge, UK.

Bailer, J. A., and C. J. Portier. 1990. A note on fitting one-compartment models: non-linear least versus linear least squares using transformed data. Journal of Applied Toxicology 10:303–306.

Benedetti, A., and G. Sebastiani. 1996. Determination of potentially mineralizable nitrogen in agricultural soil. Biology and Fertility of Soils 21:114–120.

Berg, B., and G. Ekbohm. 1991. Litter mass loss rates and decomposition patterns in some needle and leaf litter types. Long-term decomposition in a Scots pine forest. VII. Canadian Journal of Botany 69:1449–1456.

Bottcher, J. 2004. Uncertainties of nonlinearly estimated parameters from incubations of soil organic matter. Journal of Plant Nutrition and Soil Science 167:293–302.

Carroll, R. J., and D. Ruppert. 1988. Transformation and weighting in regression. Chapman and Hall, New York, New York, USA.

Cook, D. R., and S. Weisberg. 1999. Applied regression including computing and graphics. John Wiley and Sons, New York, New York, USA.

Godshalk, G. L., and R. Wetzel. 1978. Decomposition of aquatic angiosperms. II. Particulate components. Aquatic Botany 5:301–327.

Harmon, M. E. 2007. Intersite fine litter decomposition experiment (LIDET). Forest Science Data Bank, Corvallis, Oregon, USA. ⟨http://andrewsforest.oregonstate.edu/data/abstract.cfm?dbcode=TD023⟩

Harmon, M. E., W. L. Silver, B. Fasth, H. Chen, I. C. Burke, W. J. Parton, S. C. Hart, and W. S. Currie, LIDET. 2009. Long-term patterns of mass loss during the decomposition of leaf and fine root litter: an intersite comparison. Global Change Biology 15:1320–1338.

Hobbie, R. K., and B. J. Roth. 2007. Intermediate physics for medicine and biology. Fourth edition. Springer, New York, New York, USA.

Hobbie, S. E. 2005. Contrasting effects of substrate and fertilizer nitrogen on the early stages of litter decomposition. Ecosystems 8:644–656.

Hobbie, S. E. 2008. Nitrogen effects on decomposition: a five-year experiment in eight temperate sites. Ecology 89:2633–2644.

Jansson, M. 1985. A comparison of detransformed logarithmic regressions and power function regressions. Geografiska Annaler Series A, Physical Geography 67:61–70.

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

NIST [National Institute of Standards and Technology]. 2008. NIST/SEMATECH e-handbook of statistical methods. ⟨http://www.itl.nist.gov/div898/handbook/⟩

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

Packard, G. C. 2009. On the use of logarithmic transformations in allometric analyses. Journal of Theoretical Biology 257: 515–518.

Packard, G. C., and T. J. Boardman. 2009. Bias in interspecific allometry: examples from morphological scaling in varanid lizards. Biological Journal of the Linnean Society 96:296–305.

R Development Core Team. 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ⟨http://www.R-project.org⟩

Smith, J. L., R. R. Schnabel, B. L. McNeal, and G. S. Campbell. 1980. Potential errors in the 1st-order model for estimating soil-nitrogen mineralization potentials. Soil Science Society of America Journal 44:996–1000.

Smith, R. J. 1993. Logarithmic transformation bias in allometry. American Journal of Physical Anthropology 90:215–228.

Talpaz, H., P. Fine, and B. Baryosef. 1981. On the estimation of N-mineralization parameters from incubation experiments. Soil Science Society of America Journal 45:993–996.

Wieder, R. K., and G. E. Lang. 1982. A critique of the analytical methods used in examining decomposition data obtained from litter bags. Ecology 63:1636–1642.

APPENDIX A

Literature review references: 498 litterbag studies (*Ecological Archives* E091-086-A1).

APPENDIX B

Error structures of real and simulated litterbag decomposition data (*Ecological Archives* E091-086-A2).

APPENDIX C

The effect of choosing an arbitrarily small value to replace zero values (*Ecological Archives* E091-086-A3).

APPENDIX D

The effect of additional increases in simulation length on estimates of decomposition rate, k (*Ecological Archives* E091-086-A4).

APPENDIX E

The effect of changing the standard deviation (SD) of the simulated data on single-pool decomposition rate, k, estimates (*Ecological Archives* E091-086-A5).

APPENDIX F

The mean single-pool decomposition rate, k, values from linear regression of log-transformed real data and nonlinear regression of untransformed real data (*Ecological Archives* E091-086-A6).

APPENDIX G

Model comparison for Hobbie (2008) litterbag decomposition data (*Ecological Archives* E091-086-A7).

SUPPLEMENT

R code for performing nonlinear regression, with data (embedded in the R code), and a short description of the program (*Ecological Archives* E091-086-S1).