JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Fit of Four Curve−Linear Models to Decay Profiles for Pest Control Substances in Soil

ROD A. HERMAN* AND PETER N. SCHERER

Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, Indiana 46268

Experiments that investigate the pattern of degradation of pest control substances in soil are often undertaken to estimate the persistence of compounds in the environment. Mathematical models are typically fit to decay data to facilitate the interpretation of the results and make predictions concerning the environmental fate of xenobiotics in soil. Four mathematical models were fit to 61 data sets to compare their performance in conforming to empirical patterns of degradation of pest control substances in soil. The use of composite residual plots allowed comparisons of the performance of the different models over many data sets. While an exponential model, estimated using nonlinear regression, fit many data sets very well, a shift-log, biexponential, and Monod equation appears superior in many cases, and systematic deviations from data sets are often less evident with the latter models. A knowledge of the patterns of bias typically exhibited by each model across many data sets may be useful for selecting models with reduced bias when fitting individual data sets.

KEYWORDS: Soil degradation; modeling, residuals

INTRODUCTION

Experiments that investigate the pattern of degradation of pest control substances in soil are often undertaken to estimate the persistence of compounds in the environment. Mathematical models are typically fit to decay data to facilitate the interpretation of the results and make predictions concerning the environmental fate of xenobiotics in soil. Models are generally selected based on mechanistic hypotheses or empirical fit to a data set. Common models include zero-order (linear) and pseudo-first-order (exponential) equations. Zero-order relationships are characterized by a constant rate of decay over the observation period, independent of the substrate concentration.

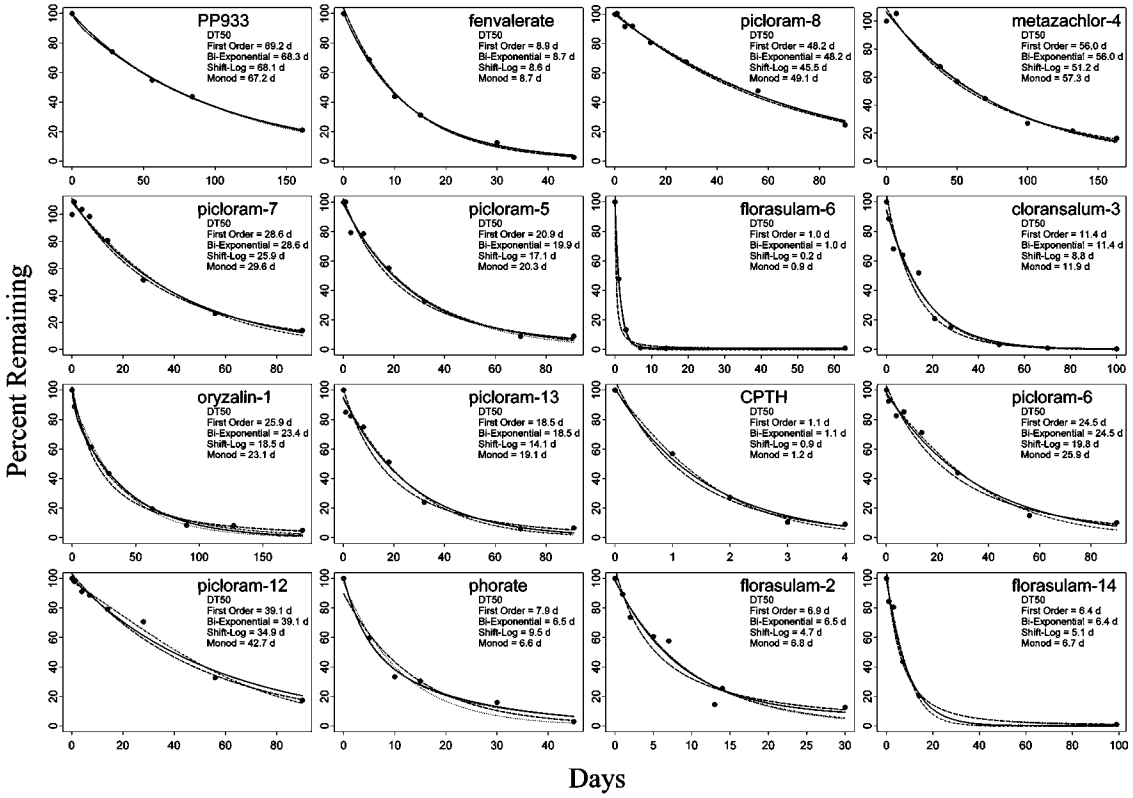
Exponential relationships exhibit a decay rate that decreases in proportion to the amount of substrate remaining. When exponential decay is observed, the logarithm of the amount remaining is linearly related to time. This relationship allows for a simple method of fitting exponential decay to experimental data using least-squares, linear regression on the transformed data. We have previously presented evidence that this method of optimizing the fit of the exponential decay model is typically less desirable than fitting the model in the natural scale using least-squares, nonlinear regression (*1*). However, in a number of cases, the nonlinear exponential approach is characterized by subtle, or sometimes more dramatic, systematic deviations from experimental data. This is especially apparent at low residues around the DT_{90} (predicted time when 90% degradation has occurred). DT_{90} values are important because they are often used to make regulatory decisions (*1*).

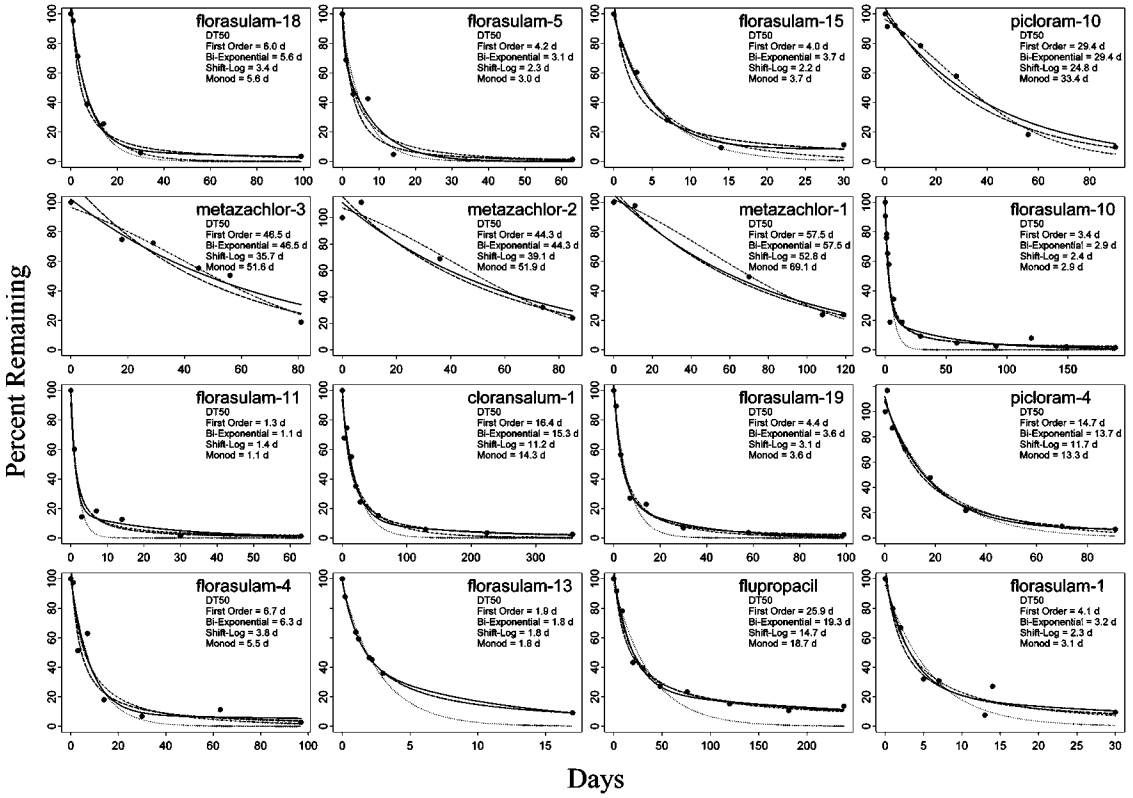
Here, we compare the fit of three additional models to that of the exponential model using nonlinear least-squares regression. We have selected these additional models for consideration based on their relative simplicity (three and four parameters) and their potential to better fit soil decay data sets that appear to be biphasic or multiphasic in nature. A four-parameter biexponential model (*2*), a three-parameter Monod equation incorporating microbe growth (*3*), and a three-parameter shift-log model (*4*, *5*) were selected for this investigation. We have not evaluated true compartment models here, as this approach typically requires experimental data quantifying partitioning into the compartments to produce meaningful results. Without this type of information, dramatically different parameter estimates can result in an equally good fit to some data sets. Also, we have also not evaluated a zero-order model here due to its simplicity and its obvious fit where appropriate (straight line in the natural scale).

Zero-order, exponential (pseudo-first-order), biexponential, Monod growth, and shift-log models have each been used to describe the degradation of pest control substances in soil, and putative mechanisms have been proposed to explain why these models may fit observed decay patterns. Zero-order decay may exist when the concentration of substrate is sufficient to saturate the metabolic capacity of microbes responsible for decomposing that particular substrate. The rate of decay is therefore mediated by the maximum capacity of these microbes and is independent of substrate concentration. Exponential decay may represent situations where the metabolic capacity of microbes responsible for degrading a particular substrate is in excess as compared to the substrate concentration. In this situation, the enzymes responsible for decay are only limited by their frequency of

* To whom correspondence should be addressed. Tel: 317-337-3551. Fax: 317-337-3255. E-mail: raherman@dow.com.

10.1021/jf060597n CCC: $33.50
Published on Web 05/13/2006


PP933
DT50
First Order = 89.2 d
Bi-Exponential = 68.3 d
Shift-Log = 68.1 d
Monod = 67.2 d
fenvalerate
DT50
First Order = 8.9 d
Bi-Exponential = 8.7 d
Shift-Log = 8.6 d
Monod = 8.7 d
picloram-8
DT50
First Order = 48.2 d
Bi-Exponential = 48.2 d
Shift-Log = 45.5 d
Monod = 49.1 d
metazachlor-4
DT50
First Order = 56.0 d
Bi-Exponential = 56.0 d
Shift-Log = 51.2 d
Monod = 57.3 d
picloram-7
DT50
First Order = 28.6 d
Bi-Exponential = 28.6 d
Shift-Log = 25.9 d
Monod = 29.6 d
picloram-5
DT50
First Order = 20.9 d
Bi-Exponential = 19.9 d
Shift-Log = 17.1 d
Monod = 20.3 d
florasulam-6
DT50
First Order = 1.0 d
Bi-Exponential = 1.0 d
Shift-Log = 0.2 d
Monod = 0.9 d
cloransalum-3
DT50
First Order = 11.4 d
Bi-Exponential = 11.4 d
Shift-Log = 8.8 d
Monod = 11.9 d
oryzalin-1
DT50
First Order = 25.9 d
Bi-Exponential = 23.4 d
Shift-Log = 18.5 d
Monod = 23.1 d
picloram-13
DT50
First Order = 18.5 d
Bi-Exponential = 18.5 d
Shift-Log = 14.1 d
Monod = 19.1 d
CPTH
DT50
First Order = 1.1 d
Bi-Exponential = 1.1 d
Shift-Log = 0.9 d
Monod = 1.2 d
picloram-6
DT50
First Order = 24.5 d
Bi-Exponential = 24.5 d
Shift-Log = 19.8 d
Monod = 25.9 d
picloram-12
DT50
First Order = 39.1 d
Bi-Exponential = 39.1 d
Shift-Log = 34.9 d
Monod = 42.7 d
phorate
DT50
First Order = 7.9 d
Bi-Exponential = 6.5 d
Shift-Log = 9.5 d
Monod = 6.6 d
florasulam-2
DT50
First Order = 6.9 d
Bi-Exponential = 6.5 d
Shift-Log = 4.7 d
Monod = 6.8 d
florasulam-14
DT50
First Order = 6.4 d
Bi-Exponential = 6.4 d
Shift-Log = 5.1 d
Monod = 6.7 d
Percent Remaining
Days


florasulam-18
DT50
First Order = 5.0 d
Bi-Exponential = 5.6 d
Shift-Log = 3.4 d
Monod = 5.6 d
florasulam-5
DT50
First Order = 4.2 d
Bi-Exponential = 3.1 d
Shift-Log = 2.3 d
Monod = 3.0 d
florasulam-15
DT50
First Order = 4.0 d
Bi-Exponential = 3.7 d
Shift-Log = 2.2 d
Monod = 3.7 d
picloram-10
DT50
First Order = 29.4 d
Bi-Exponential = 29.4 d
Shift-Log = 24.8 d
Monod = 33.4 d
metazachlor-3
DT50
First Order = 46.5 d
Bi-Exponential = 46.5 d
Shift-Log = 35.7 d
Monod = 51.6 d
metazachlor-2
DT50
First Order = 44.3 d
Bi-Exponential = 44.3 d
Shift-Log = 39.1 d
Monod = 51.9 d
metazachlor-1
DT50
First Order = 57.5 d
Bi-Exponential = 57.5 d
Shift-Log = 52.8 d
Monod = 69.1 d
florasulam-10
DT50
First Order = 3.4 d
Bi-Exponential = 2.9 d
Shift-Log = 2.4 d
Monod = 2.9 d
florasulam-11
DT50
First Order = 1.3 d
Bi-Exponential = 1.1 d
Shift-Log = 1.4 d
Monod = 1.1 d
cloransalum-1
DT50
First Order = 16.4 d
Bi-Exponential = 15.3 d
Shift-Log = 11.2 d
Monod = 14.3 d
florasulam-19
DT50
First Order = 4.4 d
Bi-Exponential = 3.6 d
Shift-Log = 3.1 d
Monod = 3.6 d
picloram-4
DT50
First Order = 14.7 d
Bi-Exponential = 13.7 d
Shift-Log = 11.7 d
Monod = 13.3 d
florasulam-4
DT50
First Order = 6.7 d
Bi-Exponential = 6.3 d
Shift-Log = 3.8 d
Monod = 5.5 d
florasulam-13
DT50
First Order = 1.9 d
Bi-Exponential = 1.8 d
Shift-Log = 1.8 d
Monod = 1.8 d
flupropacil
DT50
First Order = 25.9 d
Bi-Exponential = 19.3 d
Shift-Log = 14.7 d
Monod = 18.7 d
florasulam-1
DT50
First Order = 4.1 d
Bi-Exponential = 3.2 d
Shift-Log = 2.3 d
Monod = 3.1 d
Percent Remaining
Days

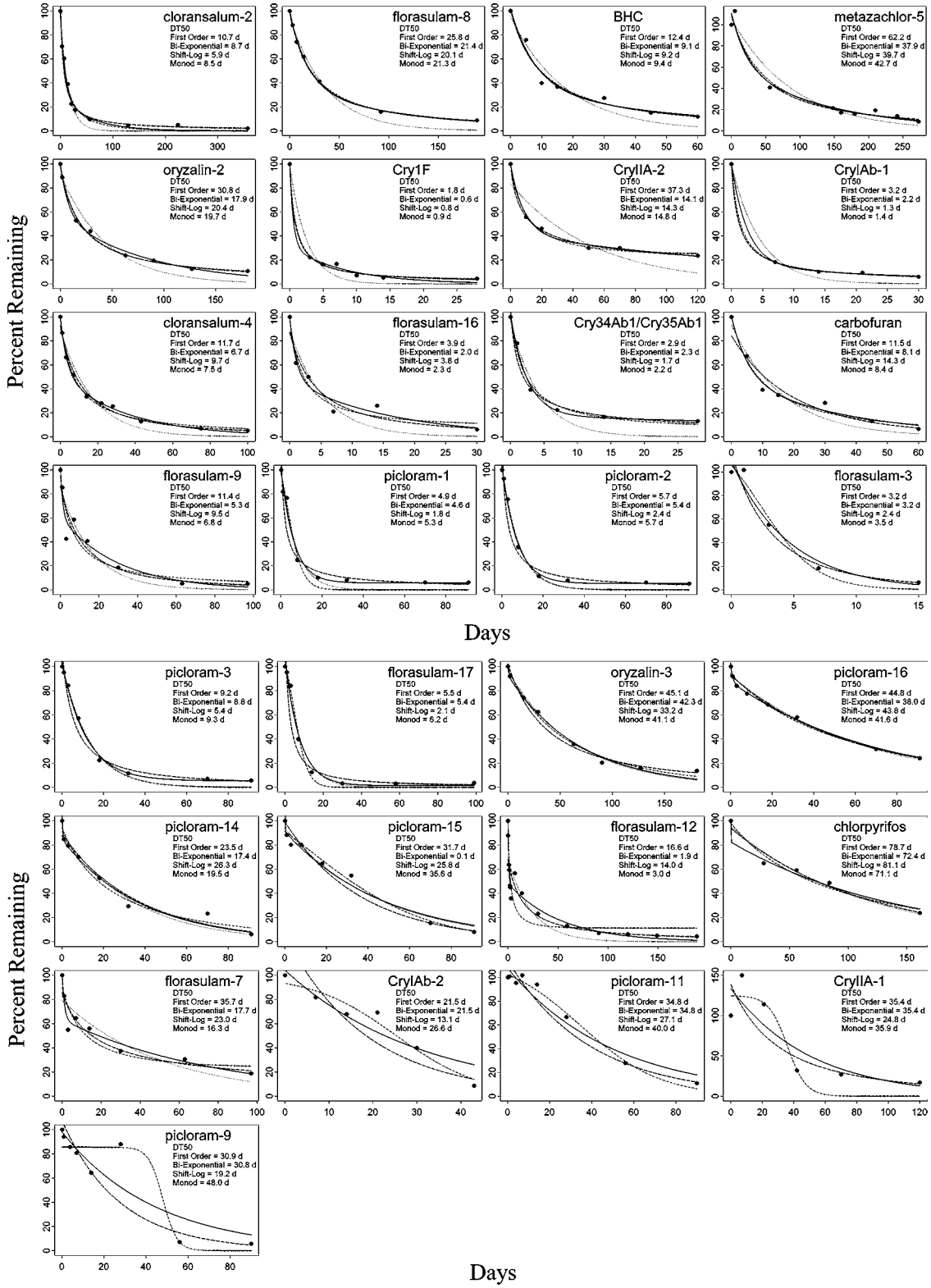


Figure 1. Fit of exponential (dotted line), shift-log (dashed line), biexponential (solid line), and Monod (dot–dash line) regression lines to soil degradation data sets. Plots are ordered subjectively based on the similarity of the models.

interaction with the substrate. Under such conditions, Michaelis−Menten enzyme kinetics predict a first-order decay pattern (*6*). The biexponential model allows for two distinct compartments to exist in soil where the substrate is immediately and statically sequestered into each compartment and where two different exponential decay rates operate. The Monod growth model also assumes Michaelis−Menten enzyme kinetics but incorporates the population growth of microbial decomposers. Finally, a mechanism for the shift-log model has been postulated based on spatial variability within soils (*4*).

Our approach was to compare the empirical fit of the curve−linear models described above for their ability to conform to experimental data from soil degradation experiments. Previously, we compared a subset of analogous models using an abbreviated data set and used the coefficient of determination as the only measure for goodness of fit (*5*). Here, we used an expanded

array of data sets, and our primary diagnostic tools for comparing the models were composite residual plots constructed on normalized axes (*1*). This tool highlights typical patterns of systematic deviations from the models and the general error structure. We also expanded this analysis by looking at time residuals to assess deviations between observations and predictions around point estimates of decay (e.g., half-lives).

MATERIALS AND METHODS

Sixty-one data sets were identified where results were amenable to analysis (*1*). Data were normalized to the initial observed residue (100%). First-order (% remaining $= S_0 \times e^{\text{slope} \times \text{time}}$), biexponential (% remaining $= S_1 \times e^{\text{slope}_1 \times \text{time}} + S_2 \times e^{\text{slope}_2 \times \text{time}}$), Monod growth [% remaining $= (P \times S_0 \times e^{P \times \text{time}})/(P + Q \times S_0 - Q \times S_0 \times e^{P \times \text{time}})$], and shift-log {log10(% remaining) = slope × [log10(time + k)] + S} models were fit to each data set using nonlinear, least-squares regression (*7*). The form of the shift-log model was chosen to facilitate convergence of the nonlinear regression routine. Nonlinear regression is an iterative process, and a solution is therefore not guaranteed. Plots for each data set were prepared showing the fit of each model (**Figure 1**).

For each model, composite plots were constructed where all residuals (deviations from the model) were plotted on normalized axes. The *y*-axis was normalized to the percent remaining substrate based on the initial observation, as described earlier. The *x*-axis (time scale) was normalized to the observed percent remaining based on the initial observation and, in a second set of plots, was normalized to the predicted percent remaining based on the predicted *y*-intercept (predicted amount of substrate at the zero time point = 100%). The percent remaining was also rescaled in the second set of plots based on the intercepts estimated by each model. Residuals were calculated by subtracting the predicted percent remaining from the observed percent remaining.

A third set of plots was constructed for each model to illustrate the pattern of time residuals. The predicted time was subtracted from the observed time for each data point and divided by the observed time. This value was multiplied by 100 to generate time residuals as a percentage of the observed value. This scaling allowed degradation patterns of greatly varying duration to be normalized. The time residuals were plotted against predicted percent decay to allow predicted point estimates of decay, such as DT_{50} (time until 50% decay) and DT_{90} values, to be compared to observed data. **Figure 2** illustrates how the three composite residual plots were constructed. Box and whisker plots were also generated to help in the comparison among models.

RESULTS AND DISCUSSION

Models. Four models were chosen for comparison based on their relative simplicity (2−4 parameters) and their ability to fit experimental data from soil degradation experiments. Relatively simple models were chosen due to the limited number of data points that typically characterize these types of data sets (**Figure 1**). The first-order model was included because it is one of the most commonly used models for fitting soil degradation data. The four-parameter biexponential model and three-parameter shift-log and Monod growth models were chosen because they are relatively simple and because these models estimate the intercept. A three-parameter biexponential model and a two-parameter shift-log model are available that each force the estimated line through the initial data point at 100% remaining (*5*). For these models, a valid analysis requires omitting the initial data point from the regression, since this data point is assigned a value of 100% and the line is forced through this point. Removing this point reduces the degrees of freedom for the analysis and is thus equivalent to adding a parameter in terms of the power of the analysis. For this reason, the simplified models offer no advantage unless the data are normalized to an estimate of the initial concentration that is independent of the initial observed concentration. This may be possible when a theoretical initial concentration is used for normalizing the data and where the observed initial observation at time zero is not set to 100% remaining.

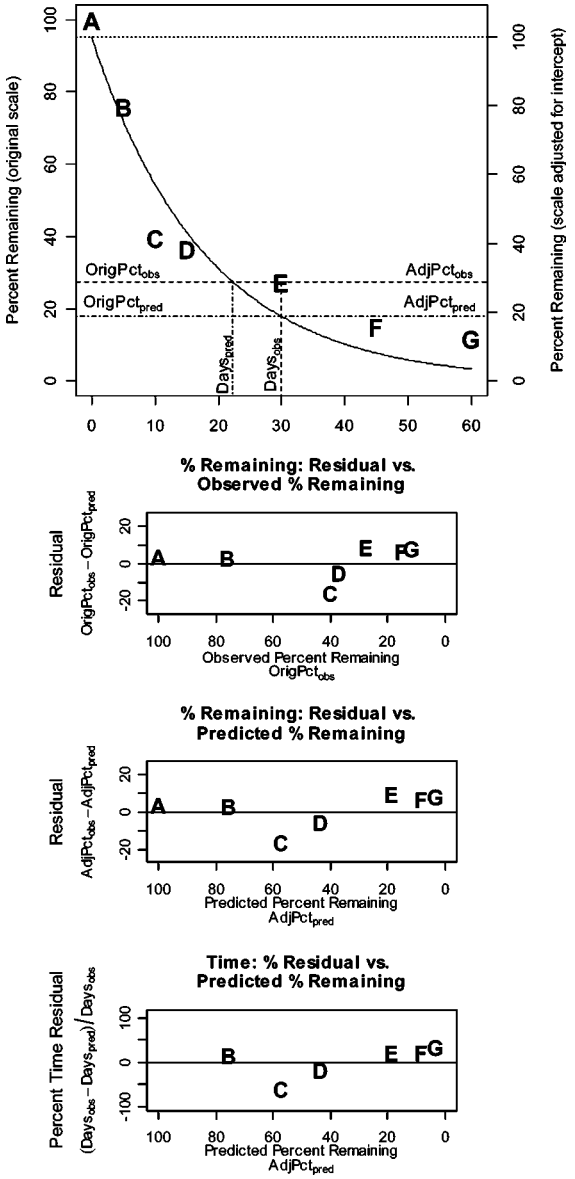


Figure 2. Mechanics of constructing composite residual plots. Panel 1 depicts the fit of nonlinear first-order model to the BHC data set, with key attributes for data point "E" noted. Panels 2–4 illustrate how the attributes depicted in panel 1 are used to construct the three different types of composite residual plots used in this paper.

Although each of these models has a putative mechanistic explanation for its fit to soil decay data sets, we chose these models based on their empirical fit to the data sets. Many mechanisms may contribute to the fit of a particular model to soil decay data, and a good fit by one of the models does not alone substantiate a particular mechanism as being responsible for the observed decay. This is evident in **Figure 1** where multiple models may fit a particular data set very well.

Model Convergence. While the convergence routine used here (S-Plus, *7*) was successful for all models and data sets, we experienced convergence problems for the biexponential, Monod growth, and shift-log models for some data sets using other nonlinear routines (e.g., SAS PROC NLIN, *8*). These problems were typically associated with data sets that were well-fit by zero-order (linear in natural scale) or first-order models. These

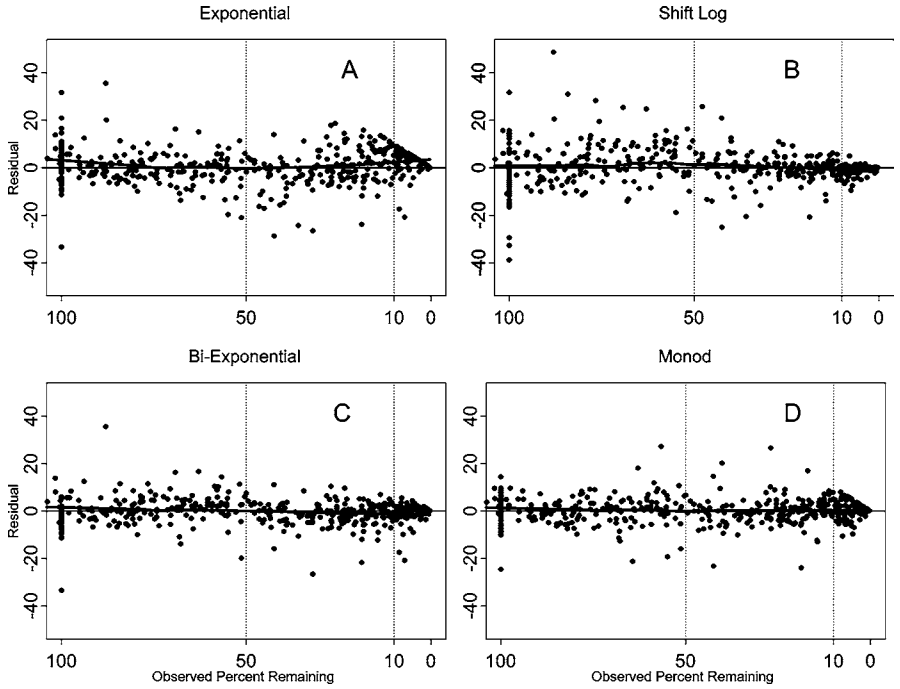


Figure 3. Residuals (observed – predicted) for indicated models plotted against the observed percent remaining.

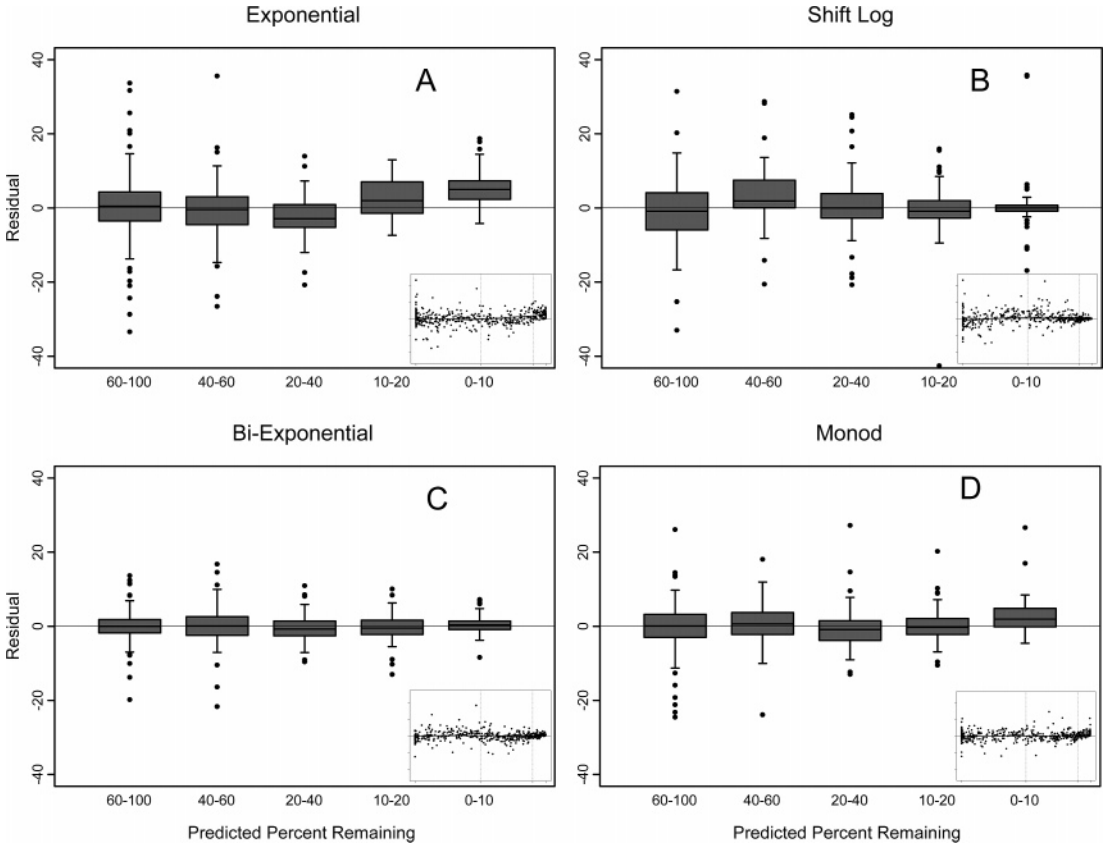


Figure 4. Residuals (observed – predicted) for indicated models (adjusted for the predicted intercept) plotted against the predicted percent remaining. Insets depict a scatter plot of the residuals, and the larger plots summarize the data into categories as indicated. The boxes encompass the middle 50% of the data (with the median dividing the box). Whiskers go to largest and smallest points within 1.5 times the interquartile range from box boundaries, and outliers are displayed as points beyond the whiskers.

convergence problems seem to be of little practical consequence because they were typically associated with data sets that are well-fit by simpler models (zero-order or first-order).

Model Fit. Figure 1 illustrates the fit of each model to each of the data sets. In general, all four of the models investigated here fit the data sets well. As noted earlier (*1*), the nonlinear, first-order model generally fits these data sets well but often appears to slightly underestimate small residues.

Composite residual plots were prepared on a normalized scale allowing the fit of each model to be evaluated across all 61 of

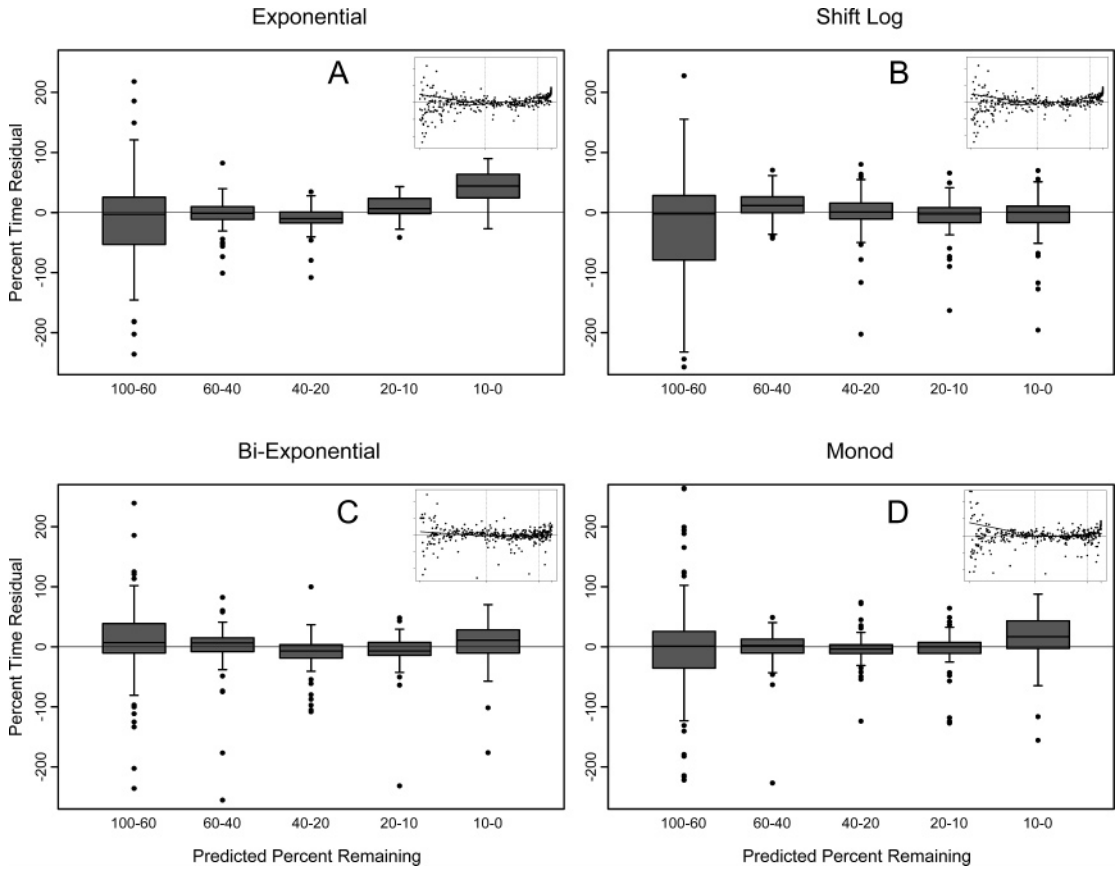


Figure 5. Time (horizontal) residuals for indicated models plotted against the predicted percent remaining. Insets depict a scatter plot of the residuals, and the larger plots summarize the data into categories as indicated. The boxes encompass the middle 50% of the data (with the median dividing the box). Whiskers go to largest and smallest points within 1.5 times the interquartile range from box boundaries, and outliers are displayed as points beyond the whiskers.

the data sets examined here (**Figures 3**−**5**). **Figure 3** depicts the residuals of the percent-remaining data vs decay, expressed as percent remaining values scaled to the initial data point (time = 0). This scaling results in a horizontal distribution of data points across the *x*-axis that is identical in each of the plots. Data points above the horizontal zero line indicate that the model predicts residues below those observed, and data points under the line indicate that the model predicts residues above those observed.

Figure 3 illustrates that each of the models appears to fit most of the data sets well and that only minor systematic deviations appear to be present for these models. There do not seem to be any obvious patterns in the error structure that indicate heteroscedasticity with the exception of the shift-log model that appears to be less variable when residues are small. This may be due to optimization of the shift-log model in the logarithmic scale, which weights small residues more heavily than would be the case in the natural scale. The four-parameter biexponential model (**Figure 3C**) and Monod model (**Figure 3D**) generally appear to better fit the data as compared to the two-parameter nonlinear, first-order model (**Figure 3A**) or the three-parameter shift-log model (**Figure 3B**). This is indicated by the reduced spread of residuals around the horizontal zero line in **Figure 3C**,**D**, as compared to **Figure 3A**,**B**. The shift-log model appears to fit smaller residues more closely as compared to the first-order model. This may be partially due to the estimation of the shift-log model in a logarithmic scale with respect to residue, since this transformation weights smaller residues more heavily than the other models.

While **Figure 3** is quite useful for comparing the fit of these models due to the equal horizontal distribution of data points, point estimates of decay such as DT_{50} or DT_{90} values are often used to estimate persistence. Normally, these values are calculated based on the intercept estimated by each model and not the value at the zero time point. For this reason, the residuals were also plotted vs the predicted percent remaining residue (**Figure 4**, insets). In addition, the residuals were rescaled to the intercept predicted by each model. Plotting residuals vs the predicted percent remaining allows for an assessment of model performance around given point estimate of decay that may be of interest. For example, persistence is often expressed as a half-life for decay that follows a first-order degradation pattern or as DT_{50} or DT_{90} values for this and other models. Accurate DT_{90} estimates are important because they are often used to make regulatory decisions relative to environmental persistence. Box and whisker plots were used to summarize the data and aid in comparing the models (**Figure 4**).

The slight underestimation of residues around the DT_{90} is readily apparent for the first-order model in **Figure 4A**. The shift-log and biexponential models both appear to estimate small residues very well (**Figure 4B**,**C**). However, the shift-log model may slightly underestimate residues around the DT_{50}. The biexponential model appears to fit most data sets very well across the entire range of residues.

Because of the varying slope of the decay curves as degradation progresses for all four models investigated here, the effect of mis-specifying residues around different point estimates of decay (e.g., half-lives) has a varying effect on the

magnitude of mis-specification of these point estimates. For example, the same magnitude of error specifying residue on a flat portion of a decay curve, which may occur when residues are small, will more greatly affect mis-specification of a DT_{90} than will the same magnitude of error on a steep portion of the curve that may occur around the DT_{50}. This is a result of larger changes in residue that occur over shorter periods of time earlier in the decay process. For this reason, a third set of plots (**Figure 5**) were constructed showing the deviation between observed data points and predicted time (horizontal residuals). These time residuals are plotted vs predicted percent remaining to allow assessment around point estimates of decay as predicted by the models. Plots illustrate the percent deviation between the point estimates of decay and the observed data. Again, box and whisker plots were used to facilitate model comparisons.

All models show the greatest spread of time residuals early in the decay process (**Figure 5**, insets). The time residuals are expressed as a percentage of the observed time to allow many data sets with greatly varying time scales to be combined in one plot. This representation of the time residuals can also result in small absolute time residuals being expressed as a large percent difference early in the decay process. For example, a point estimate that deviates from an observation by 1 day would represent 100% time residual for an observation made at 1 day but would represent a 10% time residual for an observation made at 10 days (also time residuals cannot be calculated for time zero). This is of little consequence since, in practice, most point estimates of decay are reported at or above 50% decay ($\geq DT_{50}$). For this reason, we focus our discussion on the middle and second halves of the time residual plots (four rightmost boxes in each plot in **Figure 5**).

The first-order model may slightly overestimate point estimates around the DT_{50} in some cases but certainly seems to consistently underestimate point estimates around the DT_{90} (**Figure 5A**). The Monod and biexponential models do not appear to display systematic bias in estimating persistence around the DT_{50} (**Figure 5C**,**D**) while the shift-log model appears to often slightly underestimate DT_{50} values (**Figure 5B**). However, the biexponential and Monod models may display a small bias in underestimating small residues.

In conclusion, the nonlinear, first-order model, shift-log model, and biexponential model all appear to fit the pattern of degradation of pest control substances in soil very well. On the basis of the three different composite residual plots, the first-order model often appears to slightly underestimate small residues, and this may result in systematic underestimation of DT_{90} values (**Figures 3A**, **4A**, and **5A**). In general, the more complex models appear to better fit a wider array of data sets.

Because of the limited number of data points in many individual data sets (**Figure 1**), overparametrization of models is a concern for these types of experiments. However, use of models that show systematic deviations from many data sets can result in biased conclusions. Use of first-order kinetics to model the decay of pest control substances in soil is widely accepted, easy to interpret, expected by many regulatory agencies, and required for generating rate constants for input into common environmental computer models. The first-order model also has only two parameters, making it suitable for very small data sets. However, we have illustrated here that the shift-log, biexponential, and Monod models often conform to data sets better than the first-order model (**Figures 3**−**5**). Each of the later models appears to offer similar abilities to estimate DT_{50} and DT_{90} values with limited error and with little bias but at the cost of adding one or two parameters to the model.

In the absence of knowledge concerning the mechanism of decay, the decision as to what model is appropriate for a given data set should be dependent on model fit, lack of bias, simplicity, regulatory requirements, and need to provide rate constants for environmental models. We have illustrated a common bias for the nonlinear, first-order model, and this bias can be evaluated for a specific data set to determine if use of a more complex model, such as one of those described here, is warranted. A knowledge of the patterns of bias typically exhibited by each model across many data sets may be useful for selecting models with reduced bias for use with individual data sets.

ACKNOWLEDGMENT

We thank Jim Knuteson, Kathryn Clayton, and John Cuffe of Dow AgroSciences for reviewing a draft of this manuscript.

LITERATURE CITED

(1) Herman, R. A.; Scherer, P. N. Comparison of linear and nonlinear regression for modeling the first-order degradation of pest-control substances in soil. *J. Agric. Food Chem*. **2003**, *51*, 4722−4726.

(2) Wolt, J. D.; Nelson, H. P., Jr.; Cleveland, C. B.; van Wesenkeeck, I. J. Biodegradation kinetics for pesticide exposure assessment. *Rev. Environ. Contam. Toxicol*. **2000**, *169*, 123−164.

(3) Sulfita, J. M.; Smolenski, W. J.; Robinson, J. A. Alternative nonliniear model for estimating second-order rate coefficients for biodegradation. *Appl. Environ. Microbiol.* **1987**, *53*, 1064−1068.

(4) Gustafson, D. I.; Holden, L. R. Nonlinear pesticide dissipation in soil: A new model based on spatial variability. *Environ. Sci. Technol*. **1990**, *24*, 1032−1038.

(5) Herman, R. A.; Scherer, P. N.; Wolt, J. D. Rapid degradation of a binary, PS149B1, δ-endotoxin of *Bacillus thuring*iensis in soil, and a novel mathematical model for fitting curve-linear decay. *Environ. Entomol.* **2002**, *31*, 208−214.

(6) Rawn, J. D. The Michaelis−Menten rate law governs enzymatic reactions. In *Biochemistry*; Neil Patterson Publishers: Burlington, NC, 1989; pp 167−173.

(7) *S-Plus 2000, Modern Statistics and Advanced Graphics*; Mathsoft, Inc.: Seattle, WA, 1999.

(8) SAS. *SAS/STAT User's Guide*, version 8; SAS Institute: Cary, NC, 1999.

Received for review March 1, 2006. Revised manuscript received April 17, 2006. Accepted April 18, 2006.

JF060597N