Comparison of Linear and Nonlinear Regression for Modeling the First-Order Degradation of Pest-Control Substances in Soil

ROD A. HERMAN* AND PETER N. SCHERER

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

First-order kinetic models are often used to profile the degradation of pest-control compounds in soil. This approach is based on enzyme theory and is often favored due to its simplicity and its requirement by regulatory agencies. Here, linear and nonlinear regression approaches to modeling first-order decay are compared. Composite residual plots of many soil degradation data sets are presented on a normalized scale. These plots illustrate the general error structure for the data and are useful for detecting common mis-specifications of the models. Results indicate that a nonlinear regression approach to modeling first-order decay of compounds in soil more accurately describes most data sets when compared with a linear approach. Specifically, the observed error structure does not support the broad use of a logarithmic transformation to stabilize the variance. In addition, models generated using the linear approach generally exhibit more dramatic systematic deviations from the observations as compared with models generated using the nonlinear approach. The analysis methods described here may be useful for comparing alternative models in this and other research areas.

KEYWORDS: Soil degradation; first-order kinetics; modeling; residuals

INTRODUCTION

Determining how rapidly a substrate degrades is a common objective of many scientific studies. The degradation rate for a pest-control substance is often considered when the safety of a potential product is assessed. For example, the in vitro digestibility of insecticidal crystal proteins derived from *Bacillus thuringiensis* has been used, in part, to assess the human safety of transgenic plants that express these proteins (*1*), and the degradation of pest-control substances in soil has been used to help assess potential exposure in the environment. However, there is often significant error associated with measuring the concentration of materials as they decay. For this reason, mathematical models are usually fit to the data to allow estimates of the degradation rate to be made. In addition to fitting a smooth response to potentially variable observations, such models allow specific point estimates of decay such as a DT_{50} (time until 50% decay) or DT_{90} (time until 90% decay) to be calculated. Point estimates of decay are useful for comparing the degradation rates among different materials and may help determine potential exposure.

Here we explore two common approaches to modeling the decay of pest-control substances in soil based on first-order kinetics. First-order decay is often the starting point for modeling the degradation of pest-control substances in soil. This is based on the hypothesis that the degradation is mediated by enzymes present in the microbial population of the soil and that these enzymes act much as they would in solution (*2*). It is also assumed that these enzymes are at much higher concentration than the pest-control substance. Under these conditions, the Michaelis−Menten rate law would predict a first-order decay pattern (*3*). More sophisticated mechanistic models address potentially more complex kinetics, such as soil adsorption and desorption, where first-order degradation is constrained by the availability of the pest-control substance to soil microbes and/or where two or more different first-order decay rates occur in two or more different soil compartments. A convenient attribute of substances that decay according to simple first-order kinetics is that the rate of degradation is proportional to the amount of substrate available, allowing a specific type of DT_{50}, a half-life, to describe the degradation pattern over the entire duration of decay. For this reason, regulatory agencies often favor this approach even when more complex mechanistic models fit the data more closely (*4*).

The traditional method of modeling first-order decay is to logarithmically transform the percent remaining compound and to estimate the relationship between the transformed percent remaining compound and the degradation interval (time) using least-squares, linear regression. An alternative method, made commonly accessible by advances in computer technology, is to determine the regression line in the natural scale using least-squares, nonlinear regression. An advantage of the linear regression approach is that a mathematical solution is always possible. Nonlinear regression is an iterative process and, in some cases, a solution may be difficult or impossible to obtain. If a degradation data set follows a perfect first-order decay pattern, then the two approaches provide the same mathematical solution. Therefore, the difference between the approaches is

* Author to whom correspondence should be addressed [telephone (317) 337-3551; fax (317) 337-3255; e-mail raherman@dow.com].

Degradation of Pest-Control Substances in Soil

J. Agric. Food Chem., Vol. 51, No. 16, 2003 **4723**

how they handle deviations in the observed percent remaining compound from the predictions made by the model (residuals). The linear approach minimizes the squared distances between the logarithmically transformed observations of percent remaining compound and the estimates based on the regression line, and the nonlinear approach minimizes the squared distances between the percent remaining compound and the estimates in the natural scale.

Several hypothetical arguments have been raised with regard to choosing linear or nonlinear approaches to estimating the first-order decay of pest-control substances in soil. An assumption made when using either approach is that the error structure around the observed percent remaining compound is homogeneous in the scale under which the regression is conducted. It is also assumed for both methods that the time variable has no significant error associated with it. Depending on the analytical technique used to measure the percent remaining compound, one might conjecture that higher observations of percent remaining compound are associated with higher absolute error. For instance, measuring an initial percent remaining compound of 2.0 mg/g of soil may be associated with greater error on an absolute scale than measuring 0.002 mg/g at a later time point, simply because the latter measurement is constrained by the zero asymptote. Logarithmically transforming data, when the error is proportional to the mean, is a common method to reduce heterogeneous error (*5*). Under such conditions, one could argue that the linear approach where the percent remaining compound is logarithmically transformed is most appropriate, because it hypothetically results in a more homogeneous error pattern.

Conversely, one might argue that the error structure is homogeneous in the natural scale and that logarithmically transforming the data inappropriately weights the lower observations of the percent remaining compound and later time points more heavily than the initial ones. For practical reasons, few studies are done with sufficient replication to empirically determine the preferred approach. If the error pattern for a particular study were known, weighted, least-squares regression could be used to directly correct for a nonhomogeneous error pattern.

Here we have compiled empirical data to investigate the error structure around the measurements of percent remaining compound made for pest-control substances in soil decay studies. We have generated composite plots in which the residuals of the percent remaining measurements for many data sets are plotted against a normalized decomposition scale. We have used these plots to investigate the behavior of linear and nonlinear, first-order kinetic models across many data sets and to make recommendations for their use.

MATERIALS AND METHODS

Sixty-one degradation data sets (*6–20*) were compiled where results were conducive to reanalysis (tabular form). Soil decay studies with synthetic pest-control agents and proteinaceous insect-control materials were selected from the literature. Additional results were obtained from unpublished studies conducted at Dow AgroSciences (Indianapolis, IN). For the purpose of standardization, the initial responses were set to 100%, and the remaining response values were scaled as relative percent of the initial values. This had no impact on the form of the relationship between the percent remaining compound (response) and time (covariate) or on modeling results, other than affecting the magnitude of coefficients.

Each data set was analyzed using both a first-order, least-squares, linear model and a nonlinear model (*21*). For the linear model, the natural logarithm of the percent remaining was regressed against time using linear, least-squares regression [ln(% remaining) = slope × time + intercept]. For the nonlinear model (% remaining = intercept ×

$e^{slope \times time}$), the percent remaining was regressed against time using nonlinear, least-squares regression. The latter model was used to match the form of the linear model, where the intercept is estimated and the asymptote is set to zero.

Residual plots were generated by graphing the deviation in the observed percent remaining from the estimate generated by each model. To standardize multiple data sets with greatly varying time scales on a single plot, the *x*-axis was scaled to the observed percent remaining in one set of plots and the predicted percent remaining in another set of plots. In addition to normalizing the *x*-axis scale, this allowed the performance of the models to be assessed around specific observed and predicted levels of decay such as 50% remaining or 10% remaining. The *x*-axis on the residual plots was truncated at 100% remaining for readability even though a few of data sets actually had observations above the initial observed value (>100% recovery, **Figure 1**). The residuals for plots versus predicted percent remaining were calculated from data scaled to the *y*-intercept for each model, because this is the scale generally used to calculate point estimates of decay such as half-lives. The *y*-axis for the latter plots was truncated at +200% even though a few data points were greater than this in the case of the linear model. This truncation allowed the scale of the plots to be expanded sufficiently to interpret the vast majority of the residuals. Models that fit the data well produce residual plots in which the data are uniformly distributed around the horizontal line at zero and which show minimal scatter around this line (**Figure 2**).

RESULTS AND DISCUSSION

Figure 1 illustrates the fit of both the linear and nonlinear, first-order models to each of the 61 soil degradation data sets evaluated here (in the natural scale). The plots in **Figure 1** are ordered on the basis of the agreement of the two models (percent disagreement of estimated half-lives). It should be noted that 34 of the 61 data sets describe the degradation of two materials (florasulam and picloram). However, these materials generated a full spectrum of curve shapes varying from those well fit by a first-order model to those that were fit poorly by this model. Also, when a first-order fit was poor, these materials displayed patterns similar to the other materials investigated here, in terms of how each of the first-order models mis-specified the relationship between time and decay. For this reason, the predominance of these two materials in the data sets does not appear to bias the conclusions drawn here.

The nonlinear regression converged on all of the data sets providing a model solution in each case. As expected, the overall error associated with the linear model (**Figure 2A**,**C**) is greater than that observed with the nonlinear model (**Figure 2B**,**D**). This is necessarily so because these residual plots are in the natural scale, and the linear model minimizes the squared residuals in the logarithmic scale, whereas the nonlinear model minimizes the squared residuals in the natural scale. For this reason, one must take care when comparing the two models on the basis of the overall error illustrated in **Figure 2**. However, comparison of the pattern of the error across the *x*-axis within each figure, as well as systematic departures from the models, is useful for assessing the two approaches.

Inspection of the nonlinear residuals plot versus observed percent remaining (**Figure 2B**) did not indicate any systematic patterns in the error structure with the possible exception of the zero time point (100% remaining), where a greater spread of residuals may be evident. However, initial percent remaining compound measurements are often suspect, especially for field dissipation studies (*22*). For the nonlinear model, this pattern in the data is most likely an artifact of the large number of observations at the initial time point (61 data points) and not increased error at specifying the *y*-intercept. The nonlinear model may also slightly underestimate percent remaining compound
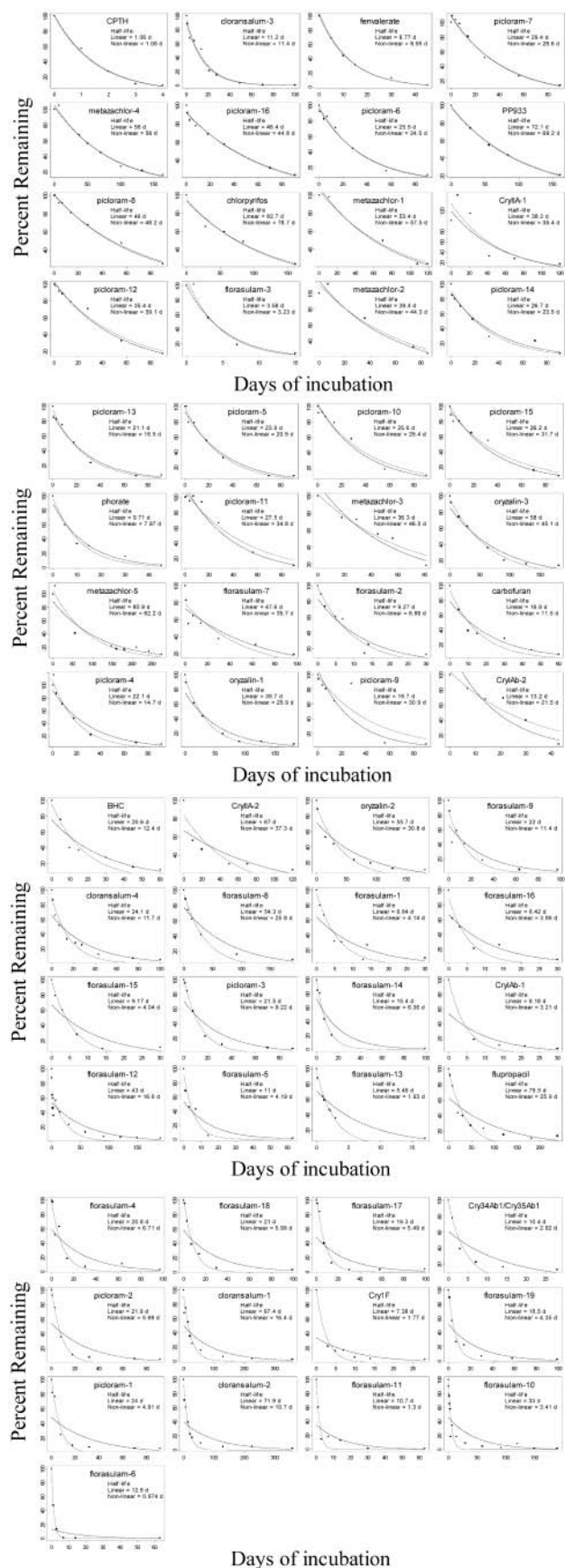
Figure 1. Fit of linear (dotted line) and nonlinear (solid line) regression lines to soil degradation data sets. Graphs are ordered based on the agreement of the half-lives estimated by each regression (as a percentage of lesser half-life).

when <10% of the compound remains, although the deviations appear to be small. The error associated with the nontransformed data does not appear to be proportional to the percent remaining compound, so, as a general rule, logarithmically transforming the data to stabilize the error does not appear to be needed.

The residuals plot for the linear regression versus observed percent remaining (**Figure 2A**) does exhibit definite nonsymmetrical patterns. Like the nonlinear model, the greatest residuals are observed at the zero time point (100% remaining). As expected, the error decreases as the percent remaining approaches 50% remaining. Because the squared residuals are minimized on logarithmically transformed percent remaining compound values, the weight of the larger percent remaining compound values are reduced more significantly compared to lower percent remaining compound values and, thus, the residuals are greater for the larger percent remaining compound measurements in the natural scale. The error for the linear model seems to be relatively stable between 50% remaining and 0% remaining. This stability may be the result of the confounding effects of experimental variability and mis-specification of the model (as described in the following paragraph).

A second observation from the residuals plot for the linear regression (**Figure 2A**) is the presence of systematic deviations from the model. For percent remaining compound measurements >50%, the model tends to underestimate the values, and for percent remaining compound measurements <50%, the model tends to overestimate the values. Such systematic deviations do not appear to be as dramatic for the nonlinear model (**Figure 2B**). It is likely that these systematic errors are caused by incorrectly specifying the model as opposed to observational error (**Figure 1**, latter plots). In studies in which the data pattern favors a two-compartment model, the values for percent remaining compound tail off more slowly than for studies in which a single-compartment, first-order decay pattern is exhibited. When the linear model is fit, a greater relative weight is given to the lower percent remaining compound values that occur at the later time points. This has the effect of decreasing the slope of the estimated decay line to a greater degree than for the nonlinear model in which all of the observations are weighted equally. Decreasing the slope of a first-order decay curve, modeled to data that fits a two-compartment pattern, will cause systematic underestimation of the larger percent remaining compound values and overestimation of the smaller percent remaining compound values. This will have the further effect of reducing the estimated y-intercept. Because measures of degradation rate are most often calculated on the basis of reductions in percent remaining compound estimates compared to the estimated y-intercept, the linear model will also systematically mis-specify these rates and point estimates such as half-lives (**Figure 2C**). Using the linear, first-order model in these situations may artificially increase such point estimates and systematically result in an overestimation of persistence.

It should be noted that an alternate explanation for an apparent biphasic decay pattern in laboratory studies may be the loss of soil viability over time (*4*). For decay that is mediated by microbes, the initial degradation rate may more accurately represent the decomposition pattern under field conditions. When the aforementioned scenario exists, systematically reducing the weight of the initial data points, as is done for the linear model, may produce misleading decay results.

Of special consideration in soil degradation studies is how well the models fit the observations around the predicted half-life and DT_{90} values for the compounds of interest. These two point estimates of decay are often used to make regulatory
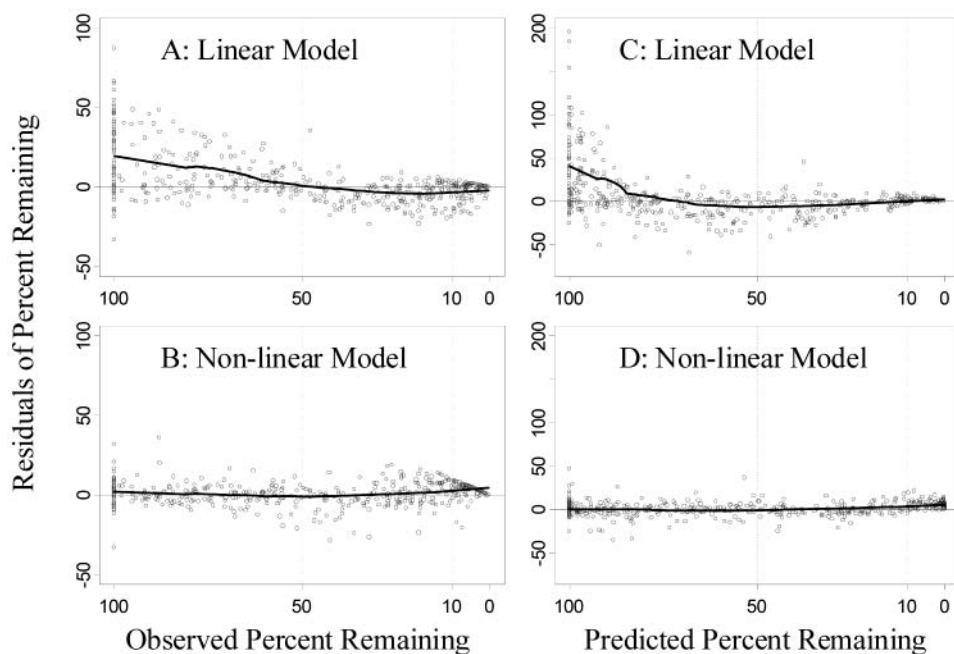
Figure 2. Residuals (observed–predicted) for the linear (A, C) and nonlinear (B, D) first-order model plotted against the observed (A, B) and predicted (C, D) percent remaining. The residuals for plots A and C are scaled on the basis of setting the initial observation to 100% remaining, and the residuals for plots C and D are scaled on the basis of setting the *y*-intercept predicted by the model to 100%. Trend lines were generated using the S-Plus supsmu function with the span set to 50% (*21*).

decisions concerning the persistence of compounds, and as such, the ability of models to fit observed values at these specific time points is especially important. Panels C and D of **Figure 2** illustrate the fit of each model versus predicted percent remaining. For these plots, the distribution of data points along the *x*-axis and the magnitude of the residuals were scaled on the basis of the *y*-intercept (initial concentration) estimated by each model. This allows evaluation of the point estimates obtained with each model.

The linear model has more error around percent remaining compound predictions at the half-life than does the nonlinear model (**Figure 2C,D**). This is not unexpected in light of the systematic error seen in the estimate of the intercept for the linear model and the diminished weight given to higher percent remaining compound values, both due to the logarithmic transformation. The linear model also tends to slightly overestimate the percent remaining compound in many cases, whereas the nonlinear model shows no apparent systematic deviation from the data in this region.

The linear model seems to generally fit well around the predicted DT_{90} values, but comparisons with the nonlinear model need to be made with caution (**Figure 2C,D**). Because the linear model often predicts longer DT_{90} values when compared with the nonlinear model, primarily due to systematic underestimation of the intercept, the percent remaining compound values observed at these later times are smaller. The percent remaining is constrained by zero asymptote and the percent remaining compound values at later time points are smaller, so absolute residuals are also necessarily smaller. These factors also contribute to a smaller number of data points in this region for the linear model, creating the illusion of less variability. The residuals around the nonlinear model are also small in this area of prediction but tend to slightly underestimate the percent remaining compound.

During our investigation, we examined the fit of two additional exponential models, each of which estimated the

asymptote rather than assigning it to zero percent remaining (not published). These models included a two-parameter model in which the intercept was assigned to 100% and the asymptote was estimated [% remaining $= a + (100 - a)e^{-kt}$] and the three-parameter model in which both the intercept and asymptote were estimated (% remaining $= a + b\,e^{-kt}$). Although these models were often excellent at fitting the data, they frequently estimated asymptotes $>10\%$, which precludes the calculation of DT_{90} values or true half-lives. This reduces the utility of these models for regulatory purposes. Because the models assume that a certain percent of the compound is infinitely stable, they are also difficult to rationalize with empirical data, which typically show systematically diminishing amounts of compound even at the latest time points.

The construction of normalized residual plots appears to be useful for evaluating the fit of a linear and a nonlinear, first-order model to the decay of pest-control substances in soil. The general error structure and systematic departures from the models are apparent in such plots. Degradation data sets are often too sparse to adequately assess the error structure of percent remaining compound measurements or to allow the specification of more complex models. Furthermore, regulatory agencies often request that a first-order model be fit to soil degradation data. The analysis of many small soil degradation data sets allows two common approaches to modeling first-order decay to be compared. The lack of an error structure showing proportionality to the mean for the nonlinear model indicates that a transformation to stabilize error is not justified as a general principle for these types of studies. Systematic departures from the linear model, and systematic misassignment of the intercept and point estimates of decay, indicate that this approach should be used with great caution.

Our analysis suggests that use of the nonlinear approach is generally preferable to the linear approach for estimating the first-order decay of pest-control substances when the error structure or a more appropriate data model is unknown. The

4726 *J. Agric. Food Chem.,* Vol. 51, No. 16, 2003

Herman and Scherer

plots depicted in **Figure 1** indicate that although many data sets are adequately represented by the nonlinear exponential model (earlier plots), this model is seen to deviate from the data in a systematic manner for many other data sets (later plots). One pattern in these plots is an underestimation of the smaller percent remaining compound values seen at advanced time points. The general pattern of mis-specification seen in the later plots of **Figure 1** may serve as an indication that, in some cases, a different model is more appropriate even if the coefficient of determination (R^2) is relatively high. For this reason, the conformation of these data sets to additional degradation models is underway in our laboratory using the approaches outlined here. The compilation of many small data sets into normalized residual plots may also provide a useful tool for evaluating competing models in other research areas.

LITERATURE CITED

(1) Astwood, J. D.; Leach, J. N.; Fuchs, R. L. Stability of food allergens to digestion in vitro. *Nat. Biotechnol.* **1996**, *14*, 1269–1273.

(2) Hemond, H. F.; Fechner, E. J. Modeling the rate of degradation. In *Chemical Fate and Transport in the Environment*; Academic Press: New York, 1994; pp 209–212.

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

(4) Wolt, J. D.; Nelson, Jr., H. P.; Cleveland, C. B.; van Wesenbeeck, I. J. Biodegradation kinetics for pesticide exposure assessment. *Rev. Environ. Contam. Toxicol.* **2001**, *169*, 123–164.

(5) Gomez, K. A.; Gomez, A. A. Logarithmic transformation. In *Statistical Procedures for Agricultural Research*; Wiley: New York, 1984; p 299.

(6) Das, A. C.; Mukherjee, D. Influence of insecticides on microbial transformation of nitrogen and phosphorus in Typic Orchragualf soil. *J. Agric. Food Chem.* **2000**, *48*, 3728–3732.

(7) Cook, W. L.; Buehrer, T. J. *The Degradation of Picloram in Four Soils at Five Application Rates*; supplemental data to U.S. Environmental Agency, subdivision N, guideline 162-1, aerobic soil metabolism; U.S. GPO: Washington, DC, 1999.

(8) Elhag, F. A.; Yule, W. N.; Marshall, W. D. Persistence and degradation of PP993 pyrethroid, fonofos, and chlorpyrifos in a Quebec cornfield's soil. *Bull. Environ. Contam. Toxicol.* **1989**, *42*, 172–176.

(9) Halliday, W. R. Environmental fate of Cry1F protein incorporated into soil. Document 7569-98-0080-AC-001 (submitted to U.S. Environmental Protection Agency for the registration of maize expressing Cry1F protein); 1998.

(10) Herman, R. A.; Collins, R. A.; Young, D. L. Degradation of microbial binary PS149B1 delta-endotoxin in a representative soil from the mid-western USA maize-growing region. Study 000365 (submitted to U.S. Environmental Protection Agency for the registration of maize expressing Cry34Ab1/Cry35Ab1 proteins); 2000.

(11) Krieger, M. S.; Merritt, D. A.; Wolt, J. D.; Patterson, V. L. Concurrent patterns of sorption–degradation for oryzalin and degradates. *J. Agric. Food Chem.* **1998**, *46*, 3292–3299.

(12) Krieger, M. S.; Yoder, R. N.; Gibson, R. Photolytic degradation of florasulam on soil and in water. *J. Agric. Food Chem.* **2000**, *48*, 3710–3717.

(13) Krieger, M. S.; Pillar, F.; Ostrander, J. A. Effect of temperature and moisture on the degradation and sorption of florasulam and 5-hyroxyflorasulam in soil. *J. Agric. Food Chem.* **2000**, *48*, 4757–4766.

(14) Rouchaud, J.; Metsue, M.; Van Himme, M.; Bulcke, R.; Gillet, J.; VanParys, L. Soil degradation of metazachlor in agronomic and vegetable crop fields. *Weed Sci.* **1992**, *40*, 149–154.

(15) Sims, S. R.; Holden, L. R. Insect bioassay for determining soil degradation of *Bacillus thuringiensis* subsp. *kurstaki* CryIA(b) protein in corn tissue. *Environ. Entomol.* **1996**, *25*, 659–664.

(16) Sims, S. R.; Ream, J. E. Soil inactivation of the *Bacillus thuringiensis* subsp. *kurstaki* CryIIA insecticidal protein within transgenic cotton tissue: laboratory microcosm and field studies. *J. Agric. Food Chem.* **1997**, *45*, 1502–1505.

(17) Spanggord, R. J.; Gordon, G. R.; Starr, R. I.; Elias, D. J. Aerobic biodegradation of [^{14}C]3-chloro-*p*-toluidine hydrochloride in loam soil. *Environ. Toxicol. Chem.* **1996**, *15*, 1664–1670.

(18) Van Wesenbeeck, I. J.; Zabik, J. M.; Wolt, J. D.; Bormett, G. A.; Roberts, D. W. Field dissipation of cloransulan-methyl at four sites in the U.S. soybean market. *J. Agric. Food Chem.* **1997**, *45*, 3299–3307.

(19) Vithala, R.; White, C. K. Aerobic metabolism of flupropacil in sandy loam soil. *Pestic. Sci.* **1996**, *47*, 235–240.

(20) Wolt, J. D.; Smith, J. K.; Sims, J. K.; Duebelbeis, D. O. Products and kinetics of cloransulan-methyl aerobic soil metabolism. *J. Agric. Food Chem.* **1996**, *44*, 324–332.

(21) S-Plus 2000; Modern Statistics and Advanced Graphics; Math-Soft, Inc., Seattle, WA, 1999.

(22) Blumhorst, M. R.; Mueller, T. C. Quality control procedures in herbicide field dissipation studies. *Weed Technol.* **1997**, *11*, 832–837.