

Article

# Error Models for the Kinetic Evaluation of Chemical Degradation Data

Johannes Ranke <sup>1,\*</sup>  and Stefan Meinecke <sup>2</sup> 

<sup>1</sup> Scientific Consultant, 79639 Grenzach-Wyhlen, Germany

<sup>2</sup> German Environment Agency (Umweltbundesamt), 06844 Dessau-Roßlau, Germany; stefan.meinecke@uba.de

\* Correspondence: johannes.ranke@jrwb.de

† Current address: Kronacher Str. 12, 79639 Grenzach-Wyhlen, Germany.

Received: 15 November 2019; Accepted: 5 December 2019; Published: 10 December 2019

**Simple Summary:** The use of a two-component error model for the evaluation of chemical degradation data is reported for the first time in this contribution.

**Abstract:** In the kinetic evaluation of chemical degradation data, degradation models are fitted to the data by varying degradation model parameters to obtain the best possible fit. Today, constant variance of the deviations of the observed data from the model is frequently assumed (error model “constant variance”). Allowing for a different variance for each observed variable (“variance by variable”) has been shown to be a useful refinement. On the other hand, experience gained in analytical chemistry shows that the absolute magnitude of the analytical error often increases with the magnitude of the observed value, which can be explained by an error component which is proportional to the true value. Therefore, kinetic evaluations of chemical degradation data using a two-component error model with a constant component (absolute error) and a component increasing with the observed values (relative error) are newly proposed here as a third possibility. In order to check which of the three error models is most adequate, they have been used in the evaluation of datasets obtained from pesticide evaluation dossiers published by the European Food Safety Authority (EFSA). For quantitative comparisons of the fits, the Akaike information criterion (AIC) was used, as the commonly used error level defined by the FORum for the Coordination of pesticide fate models and their USE(FOCUS) is based on the assumption of constant variance. A set of fitting routines was developed within the mkin software package that allow for robust fitting of all three error models. Comparisons using parent only degradation datasets, as well as datasets with the formation and decline of transformation products showed that in many cases, the two-component error model proposed here provides the most adequate description of the error structure. While it was confirmed that the variance by variable error model often provides an improved representation of the error structure in kinetic fits with metabolites, it could be shown that in many cases, the two-component error model leads to a further improvement. In addition, it can be applied to parent only fits, potentially improving the accuracy of the fit towards the end of the decline curve, where concentration levels are lower.

**Keywords:** kinetic evaluation; chemical degradation; error model; variance function

## 1. Introduction

In recent discussions, the realism and adequacy of current environmental risk assessment methods for pesticides [1], and in particular of their exposure assessment part, e.g., for surface water [2–4], have been questioned. Therefore, it will be necessary to perform more reality checks of the exposure models used in these assessments. In such checks, the uncertainties in the model predictions need to be taken

into account as well. For this, it is desirable to obtain not only point estimates, but also a description of the uncertainties of the influential input parameters like degradation rate constants or parameters for the Freundlich-type sorption to soils [5]. In the area of environmental degradation rates, previous work has brought some improvements regarding the characterization of the most likely distributions of fitted degradation parameters using frequentist [6], as well as Bayesian approaches [7–9].

The error models generally used for the concentration data in such evaluations either assume constant variance (ordinary least squares (OLS)) across all concentrations for all observed compounds and compartments (error model “constant variance”) or alternatively allow for a different variance for each observed variable [6]. For the latter error model, fitting by the iteratively reweighted least squares (IRLS) method is now widely used in regulatory risk assessments due to its availability in the kinetic evaluation software packages KinGUII [10] and CAKE [11]. In this study, we will refer to this error model using the term “variance by variable”, as it can not only be used for the case of different variances of parent compound and metabolite concentrations, but also to allow for different variances for the concentrations of the same compound in different matrices.

In contrast to the variance by variable approach, error models that have been proposed in the context of analytical chemistry cater to the common observation that the variance of the observations depends on the magnitude of the observations [12–15]. For the random errors observed in analytical chemistry, this can be explained procedurally, as some parts of the procedure may introduce errors that are largely independent of the size of the signal, like fluctuations in signal detection, while other parts may introduce variations that will result in errors that are proportional to the observed signal, like aliquoting [12]. If the former error type is prevalent, the assumption of constant variance is adequate. In the other extreme, the variance of the signal may be explained purely by a relative standard deviation (coefficient of variation).

In the field of analytical chemistry, as well as in the bioassay literature referenced therein, it has been observed that the specification of the error model often has a negligible influence on the parameters derived [13]. However, when it comes to the calculation of confidence intervals or other measures of confidence in the measured values like limits of detection, the introduction of a more appropriate error model can have a large effect on the results [14,15].

Transferring these observations to the field of degradation kinetics, it can be expected that the different error models will have a larger influence on parameter confidence intervals and parameter significance tests than on the parameter values.

In order to investigate if it is possible to derive the components of the two-component error model from the degradation data alone, without reference to calibration data, the functionality to fit a two-component error model to chemical degradation by maximum likelihood estimation has been added to the mkin software package [16] as described below.

In this paper, we describe the application of this error model to various test datasets and report how it compares to unweighted least squares fitting, as well as to the variance by variable error model proposed by Gao et al. [6]. For model comparison, the  $\chi^2$  error level proposed in the guidance of the FORum for the Coordination of pesticide fate models and their USE(FOCUS) kinetics work group [17] is not appropriate because it is based on the assumption of constant variance. As the models to be compared are not necessarily nested, the Akaike information criterion (AIC) [18] is employed for the comparisons, which is not based on the assumption of nested models.

In addition, it is discussed why methods for the direct or stepwise minimization of the complete negative log-likelihood function have been implemented for fitting the two-component error model instead of using an IRLS algorithm.

## 2. Theory and Methods

The expected value of the magnitude  $Y_i$  of the  $k$  observed variables over time  $t$  in dependence of the vector of degradation model parameters  $\beta$  is defined by the degradation model function (mean response function)  $f(i, t, \beta)$ :

$$E(Y_i|i, t, \beta) = f(i, t, \beta) \quad (1)$$

where  $i$  is an index variable, indicating the observed variable, i.e., ranging from one to  $k$ . Here, observed variables are the concentrations of parent and transformation products, optionally in different compartments. If only degradation of one compound is observed in one compartment, there is only one observed variable, and the index variable  $i$  can be ignored. If data for transformation products are observed as well, or if transfer to a different compartment is observed, the respective observed concentrations are indexed by  $i$ . The initial concentrations  $y_0$ , as far as they are to be estimated from the data, are a subset of the degradation model parameters  $\beta$ .

Supposing a normal distribution for the deviations from this deterministic degradation model, we can write the conditional distribution for an observation  $Y_j$  observed for variable  $i$  at time  $t$  as:

$$Y_j \sim \mathcal{N}(f(i, t, \beta), g(i, t, \beta, \theta)) \quad (2)$$

with  $\theta$  being additional parameters of the error model function (variance function)  $g(i, t, \beta, \theta)$ . Note that we assume that the errors for each observed variable are independent.

This notation is similar to the one used by Davidian and Carroll [19]. However, the function  $g$  is defined differently, as the standard deviation  $\sigma$  is treated here as one of the error model parameters  $\theta$ . Furthermore, the index variable  $i$  is used as an explicit argument to the degradation and error model functions, determining for which of the observed variables these functions should return the expected value, respectively the expected variance.

### 2.1. Constant Variance

Using unweighted or ordinary least squares fitting (OLS) is equivalent to using an error model:

$$g(i, t, \beta, \theta) = \sigma^2 \quad (3)$$

with the scalar standard deviation  $\sigma$  being the only component of the error model parameter vector  $\theta$ .

### 2.2. Variance by Variable

In the error model proposed by Gao et al. [6], unequal variances  $\sigma_i^2$  are allowed for the different observed variables, which means that the error model function is:

$$g(i, t, \beta, \theta) = \sigma_i^2. \quad (4)$$

Here, the vector of error model parameters consists of the different standard deviations  $\sigma_i$  for the  $k$  observed variables (types of concentrations).

### 2.3. Two-Component Error Model

Following the idea, taken from analytical chemistry, that the variance is a function of the magnitude of the observed value as approximated by  $f(i, t, \beta)$  in the fitting process, we can write:

$$g(i, t, \beta, \theta) = h(f(i, t, \beta), \theta). \quad (5)$$

In accordance with Werner et al. [12] and Wilson et al. [15], we can use:

$$g(i, t, \beta, \theta) = \sigma_{\text{low}}^2 + f(i, t, \beta)^2 \text{rsd}_{\text{high}}^2 \quad (6)$$

where  $\sigma_{\text{low}}$  represents the standard deviation at low levels and  $\text{rsd}_{\text{high}}$  the relative standard deviation at high levels. In this error model, we only have these two error model parameters, regardless of how many different observed variables are being modeled. Note that this is a simplified version of the two-component model proposed by Rocke and Lorenzato [14], which assumes the lognormal distribution for the errors at large values, which is not compatible with the above formalism.

#### 2.4. Likelihood Functions

Conventionally, the likelihood function given for random variables following a normal distribution is based on its probability density. However, in chemical kinetics, censored data are quite common, due to the presence of limits of detection (LOD) or limits of quantification (LOQ). In order to take such values into account, the likelihood for each observation could be based on an integral over the probability density.

However, in the case that no censored data are present, the likelihood of a single observation can also be specified in the usual way using the probability density of a normal distribution:

$$L_j(\beta, \theta | i, t, y_j) = \varphi \left( \frac{y_j - f(i, t, \beta)}{g(i, t, \beta, \theta)} \right). \quad (7)$$

where  $\varphi$  is the standard normal distribution, the mean response function  $f(i, t, \beta)$  is used as the estimator  $\hat{\mu}_j$  of the expected value for the observation, and the variance function  $g(i, t, \beta, \theta)$  is used as the estimator  $\hat{\sigma}_j$  of its expected variance.

As current regulatory guidance recommends to treat values below the limit of detection by replacing them with surrogate values [20], fitting degradation models to censored data based on the likelihood definition using intervals as outlined above has not been implemented yet.

The likelihood  $L$  given the complete set  $y$  of  $n$  observations  $y_j$  is obtained as the product of these likelihoods  $L_j$ , which is then maximized by minimizing the sum of the negative logarithms of these likelihoods to obtain maximum likelihood estimates for the parameters of the degradation model, as well as of the error model.

#### 2.5. Algorithms for Parameter Estimation

When using the probability density for the likelihood, we can express the likelihood function as:

$$L(\beta, \theta | i, t, y) = \prod_{j=1}^n \frac{1}{\hat{\sigma} \sqrt{2\pi}} \exp \left( -\frac{(y_j - \hat{\mu}_j)^2}{2\hat{\sigma}^2} \right) \quad (8)$$

and the log-likelihood  $\ell$  as:

$$\ell(\beta, \theta | i, t, y) = -n \ln(\hat{\sigma}) - n \ln(\sqrt{2\pi}) - \sum_{j=1}^n \frac{(y_j - \hat{\mu}_j)^2}{2\hat{\sigma}^2} \quad (9)$$

In the case that constant variance is assumed as the error model, the maximum likelihood estimates of the degradation model parameters can be obtained by minimizing the sum of squared residuals (i.e., by using OLS), because the derivatives of the log-likelihood function with respect to the model parameters only depend on the sum of squared residuals  $\sum_{j=1}^n (y_j - f(i, t, \beta))^2$ . The maximum likelihood estimator for  $\sigma$  can then be obtained as the root of this sum of squared residuals divided by the number of observations.

Alternatively, the complete likelihood function given above can be optimized directly by varying degradation model parameters and error model parameters, either simultaneously or in a stepwise procedure.

For fitting the error model with a unique variance for each observed variable, the IRLS method has been proposed [6] and implemented in the commonly used software packages KinGUII, CAKE,

and mkin (starting from version 0.9-22). After an initial OLS optimization of the degradation model parameters, the variances of the different observed variables are estimated as their root mean squared errors, and a weighted least squares fit follows. This is iterated until the variances converge up to a specified tolerance.

When the two-component error model is used in the estimation, the IRLS method is also possible. After the initial OLS step and after each subsequent weighted least squares step, the current error model components  $\sigma_{\text{low}}$  and  $\text{rsd}_{\text{high}}$  can be estimated from the residuals in different ways, for example by a nonlinear regression of the absolute residuals or their squares against the observed values or the current predictions [19].

Several variants of this approach have been implemented in previous development versions of mkin. It was found that the approach has two disadvantages in the case of the two-component error model. Firstly, fitting the error model to the residuals did not always succeed, leading to an interruption of the fitting process. Secondly, when the second, relative error component  $\text{rsd}_{\text{high}}$  was estimated to be zero, the absolute error component  $\sigma_{\text{low}}$  was not the same as the standard deviation of residuals  $\sigma$  obtained from an OLS fit to the same dataset.

Therefore, direct or stepwise minimization of the complete negative log-likelihood function was implemented as alternative approach. The consistency of the new algorithms with previous results obtained for standard datasets was ensured by using the tests integrated in the mkin package, which compare results obtained for some of the standard datasets given in the FOCUS guidance [17] and some of the synthetic datasets published previously for the purpose of software comparison [21] with known results.

## 2.6. Result Comparison Criteria

As mentioned above, it would be unreasonable to base a model comparison on the commonly used FOCUS  $\chi^2$  error level, as it is based on the assumption of a constant variance and would therefore always favor the model that has been optimized using this assumption, i.e. the OLS fit.

Therefore, a more general, likelihood based approach was needed. As the models to be compared are not necessarily nested, the very general information theory based Akaike information criterion (AIC) [18] was implemented as:

$$\text{AIC} = 2p - 2\hat{\ell} \quad (10)$$

with  $\hat{\ell}$  being the log-likelihood evaluated at the maximum likelihood estimates of the  $p$  parameters. The number  $p$  is the sum of the number of degradation model parameters  $\beta$  and the number of error model parameters  $\theta$ . A lower AIC means a preferable model. In cases where the error structure is well described by constant variance, the log-likelihood of the alternative error models (“variance by variable” or “two-component”) is approximately the same, and the term  $2p$  will lead to a higher AIC for these models because they have a higher number of parameters. Conversely, a certain improvement of the log-likelihood is necessary for these models to be seen as preferable to the simpler “constant variance” error model. Note that a lower AIC of one model does not necessarily mean that the alternatives can be ignored, especially if the differences are small. Please refer to Burnham and Anderson [22] where guidance for interpreting the AIC is given.

## 2.7. Example Datasets

In order to check these methods with data that are practically relevant, 12 soil datasets and six water sediment datasets were obtained from active substance evaluation dossiers published by the European Food Safety Authority (EFSA). These datasets and their complete evaluations using the three different error models are available as Supplementary Information. In addition, the datasets are included as example datasets in current versions of mkin [16].

### 3. Results

#### 3.1. Algorithms

Several ways to employ the three different error models in the model fitting were tested. For ordinary least squares, direct minimization of the negative log-likelihood proved to take more time than the usual least squares optimization. In addition, the optimum solution sometimes was not even found, but only a local minimum. The reason for this is to be seen in the fact that in the case of constant variance, the degradation model parameters and the single error model parameter  $\sigma$  are independent. During the optimization of the full likelihood, this leads to a singular Hessian matrix, which obviously entails a less robust optimization. Therefore, simple least squares optimization was used for the case of constant variance, and the complete likelihood was calculated in a separate step.

For fitting the two-component error model, different variants of stepwise optimization of the likelihood were tried, the first step always being an OLS fit to obtain degradation model parameters robustly. After this first step, either the complete likelihood function was optimized directly (two-step algorithm) or a separate optimization of only the error model components with fixed degradation model parameters was inserted (three-step algorithm). The four-step variant inserted another optimization of only the degradation model parameters after the error model optimization. Finally, an iterative optimization of the error model and the complete likelihood was implemented, similar to the IRLS algorithm proposed by Gao et al. [6], but always using the complete likelihood function for estimating the current error model, with degradation parameters fixed to current values.

These stepwise methods were compared with the direct minimization of the negative log-likelihood. It turned out that in order to find the fit with the highest likelihood for all twelve soil datasets and using the two-component error model, both the direct optimization of the likelihood and the three-step algorithm lined out above had to be run. Therefore, the current version of mkin (Version 0.49.5 and greater) runs both algorithms and then reports the result with the higher likelihood.

For all of the algorithms currently implemented in mkin, the constant component  $\sigma_{\text{low}}$  is equal to  $\sigma$  obtained by OLS fitting in cases where the relative component of the error model  $rsd_{\text{high}}$  is close to zero.

For the variance by variable error model, the same solutions were found for direct likelihood optimization and for the three-step optimization. As only twelve coupled datasets were tested, both algorithms are tried in the current default method as a precaution.

#### 3.2. Parent Only Data

The variance by variable model introduced by Gao et al. [6] is not different from unweighted (OLS) fitting if there is only one compound observed in one compartment and therefore only one observed variable. Therefore, in this section describing the parent only fits for the soil datasets, the two-component error model is only compared to OLS fitting.

Table 1 lists the FOCUS  $\chi^2$  error levels (goodness-of-fit measure proposed in the FOCUS guidance), as well as the AIC values of these fits. The FOCUS  $\chi^2$  error level was either equal or larger in the case of the two-component error model, because it was based on the assumption of a constant variance. In contrast, the AIC of the fits using the two-component error model was lower in the case of FOCUS Datasets A, B, and D.

**Table 1.** Error model parameters obtained for simple first order (SFO) fits to the FOCUS datasets (see text).

Dataset	OLS			Two-Component			
	$\sigma$	$\chi^2$ (%)	AIC	$\sigma_{\text{low}}$	$rsd_{\text{high}}$	$\chi^2$ (%)	AIC
FOCUS A	5.27	8.39	55.3	0.00	0.09	13.77	37.5
FOCUS B	1.96	4.46	39.5	0.32	0.04	4.89	31.0
FOCUS C	4.67	15.85	59.3	4.67	0.00	15.85	61.3
FOCUS D	3.40	6.45	101.1	0.00	0.07	6.59	46.8

In order to give a more complete picture of the effect of using the two-component error model for parent decline data, Table 2 shows a comparison of the AIC values obtained for the twelve real-world soil datasets, with simple exponential decline (simple first order (SFO)) and bi-exponential decline (dual first order in parallel (DFOP)), each fitted with unweighted least squares (OLS) and the two-component error model.

**Table 2.** Overview of AIC values (lowest in bold) obtained for simple first order (SFO) and dual first order in parallel (DFOP) (bi-exponential decline) fits to simple soil datasets.

Dataset	SFO		DFOP	
	OLS	Two-Component	OLS	Two-Component
Soil 1	131.2	133.2	<b>107.9</b>	109.9
Soil 2	137.9	138.1	<b>80.3</b>	82.3
Soil 3	78.8	80.8	67.0	<b>65.9</b>
Soil 4	82.0	84.0	<b>75.0</b>	77.0
Soil 5	<b>172.4</b>	174.4	176.0	178.0
Soil 6	128.3	122.9	<b>79.0</b>	81.0
Soil 7	127.3	129.3	89.8	<b>88.7</b>
Soil 8	93.1	93.2	64.8	<b>55.7</b>
Soil 9	137.4	134.9	<b>109.0</b>	111.0
Soil 10	111.3	110.7	76.6	<b>75.3</b>
Soil 11	<b>125.5</b>	127.5	129.5	131.5
Soil 12	82.0	82.0	<b>65.9</b>	66.9

For SFO fits, the two-component error model offered a lower AIC for Soil Datasets 6, 9, and 10. In all of these cases, an even lower AIC can be obtained when the biphasic DFOP model was used. The DFOP offered a lower AIC also in the case of the other datasets, with the exception of Soil Datasets 5 and 10, which apparently followed a simple exponential decline, without a two-component error structure.

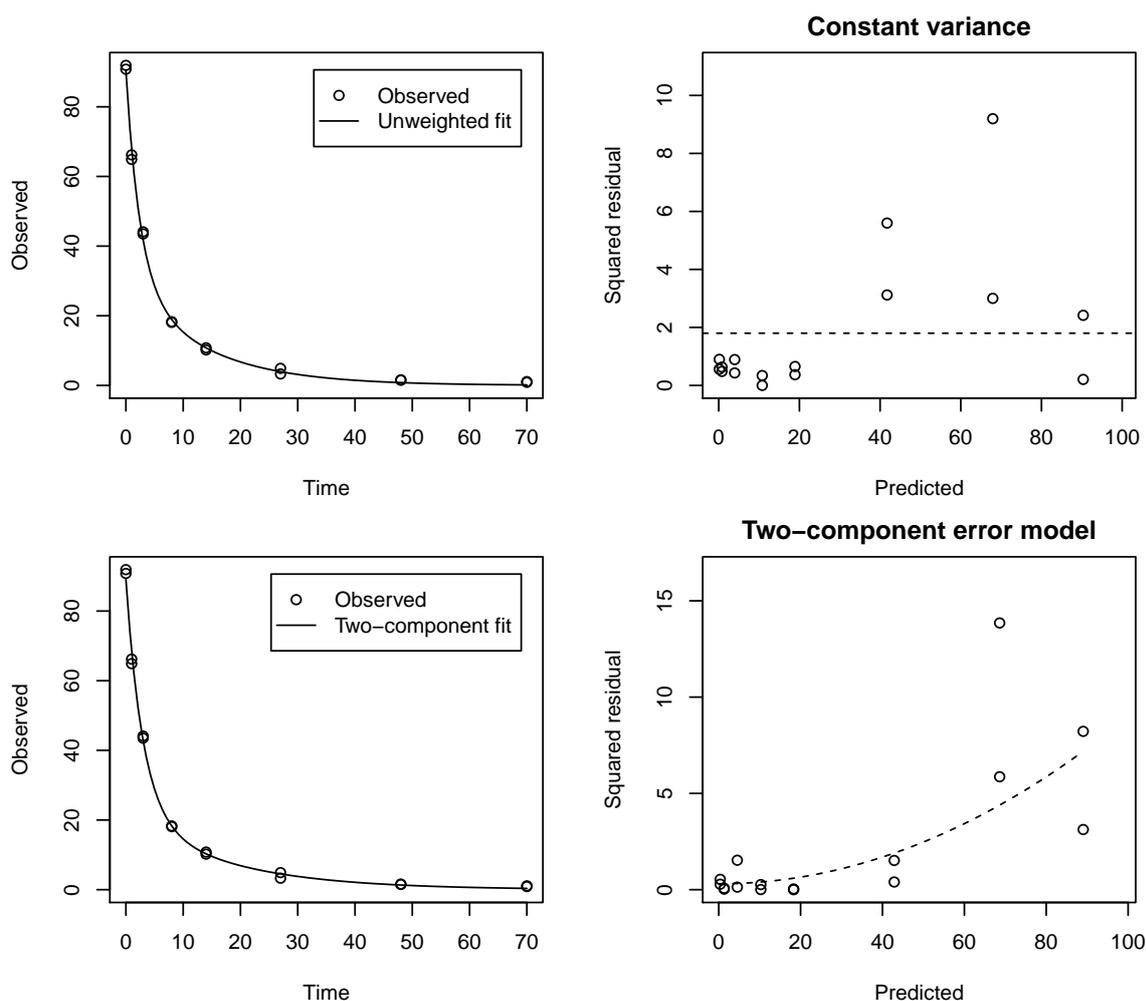
Of the ten datasets best described by the DFOP model, the two-component error model yielded an improved AIC in four cases (Soil Datasets 3, 7, 8, and 10). In all of these cases, the rate constant obtained for the slower phase using the two-component error model was either equal to or lower than the one obtained by the unweighted least squares fit (see the Supplementary Information for details).

A graphical illustration of the fits and the error structures for Soil Dataset 8, where the improvement of the AIC by using the two-component error model was most pronounced, is shown in Figure 1. The plots of the squared residuals against predicted values in the two panels to the right illustrate that there was indeed a trend in squared residuals to increase with predicted values and that this trend was adequately reflected by the two-component error model.

### 3.3. Coupled Fits

Similar observations could be made when simultaneous evaluations of parent decline and metabolite formation and decline were made with a coupled model. For each of the eleven soil datasets featuring one or more metabolites, a parent decline model (either SFO or DFOP) was selected for the parent compound, based on the parent only fits discussed above and coupled to simple first-order (SFO) degradation of the metabolite. The selected coupled models were then fitted using OLS fitting, the variance by variable model, and the two-component error model. The AIC values obtained in these coupled fits are listed in Table 3.

Only in two cases, ordinary least squares fitting provided the lowest AIC value. The variance by variable error model led to the lowest AIC for four of the datasets. In the remaining five cases, the two-component error model gave the lowest AIC, indicating that the error structure was most adequately represented by this error model.



**Figure 1.** Comparison of bi-exponential (DFOP) kinetic fits assuming constant variance (top panels) and the two-component error model (bottom panels) for Soil Dataset 8. The fitted error models are shown in the plots of the squared residuals to the right as dashed lines.

**Table 3.** Overview of AIC values (lowest in bold) obtained for the coupled fits to soil datasets with metabolites.

Dataset	OLS	Variance by Variable	Two-Component
Soil 1	252.0	242.9	<b>231.5</b>
Soil 2	<b>207.5</b>	210.9	209.5
Soil 3	110.2	108.0	<b>104.0</b>
Soil 4	125.7	<b>124.4</b>	125.4
Soil 5	266.9	<b>201.3</b>	210.2
Soil 6	<b>138.6</b>	139.8	139.0
Soil 7	156.7	155.6	<b>150.5</b>
Soil 8	106.8	105.5	<b>99.6</b>
Soil 9	181.5	<b>165.8</b>	178.7
Soil 10	130.5	129.0	<b>123.4</b>
Soil 11	268.4	<b>216.7</b>	238.5

Figure 2 shows the three fits including the three error models for Soil Dataset 3, which was one of the datasets for which the two-component error model provided the lowest AIC. The panels to the right showed that the level of squared residuals was generally lower for metabolite A1, which was the reason why the variance by variable error model gave a lower AIC than the OLS fit assuming the same variance for both compounds. However, it was also visible that there was an increasing trend of the

squared residuals with increasing fitted values, which can be described by the two-component error model. These plots suggested that the squared residuals for metabolite A1 were lower than for the parent compound because the residue levels of A1 were lower than the parent residue levels. In this case, it was more efficient to assume that the error depended on the magnitude of the observed value than to assume that it depended on the type of compound observed.

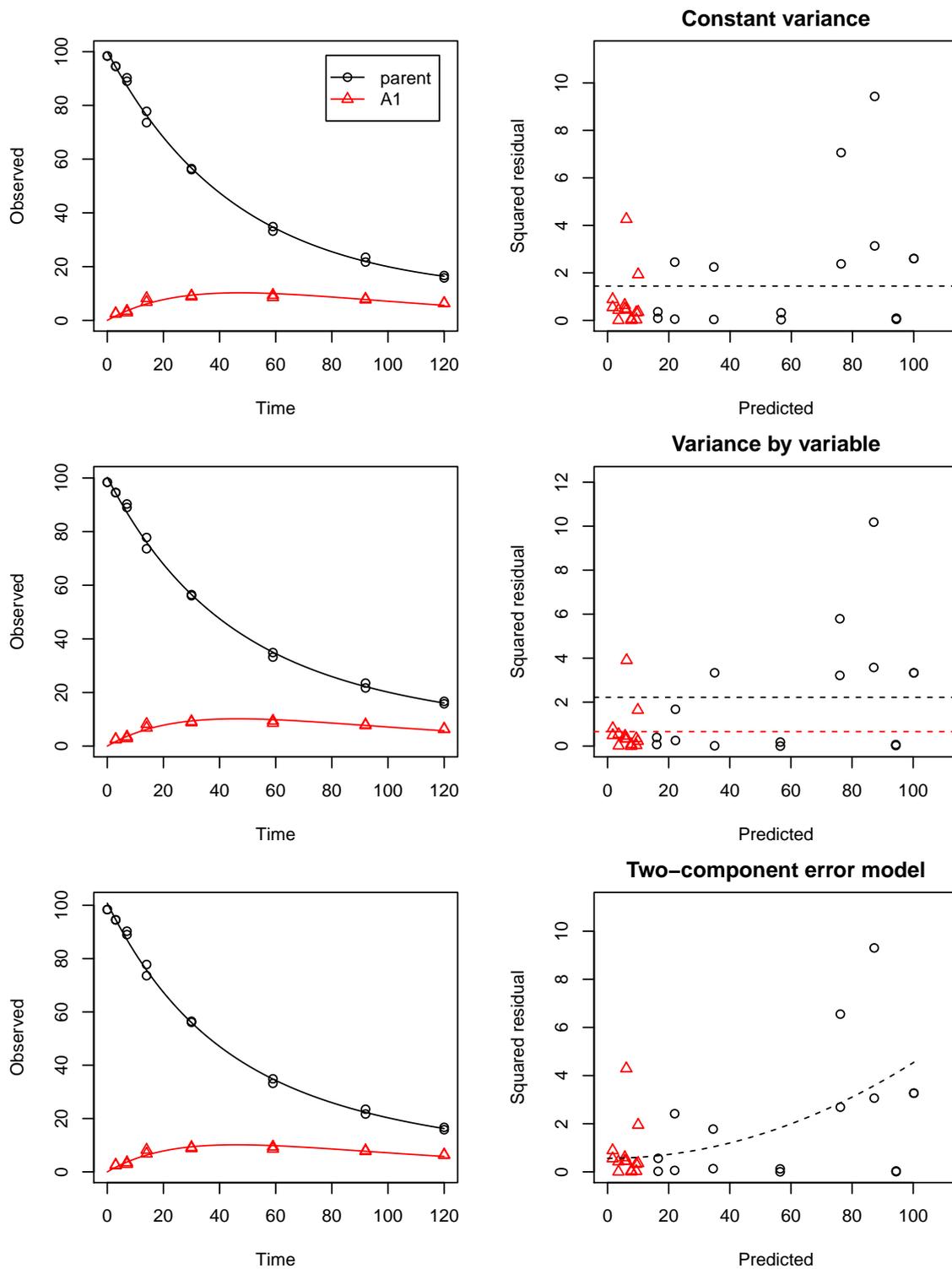


Figure 2. Comparison of three different error models for Soil Dataset 3.

The corresponding degradation model parameters with the  $p$ -values for the  $t$ -test for significant difference from zero are given in Table 4. While the value for the degradation rate constant  $k_{A1}$  for metabolite A1 was nearly the same for all three fits, an interesting observation could be made for the rate constant  $k_2$  for the second phase of the biphasic decline of the parent compound. While in the OLS fit, it had a very low value of  $1.2 \times 10^{-5}$  per day and would not be deemed significantly different from zero, it took on an appreciable value of about 0.007 per day when the variance by variable error model was used. While the  $p$ -value was still far from the usual significance trigger of 0.05 using variance by variable, it dropped to 0.04 if the two-component error model was used, and the estimate for  $k_2$  rose to around 0.008 per day.

This suggests that where the error structure of parent and metabolites was well described by the two-component error model, even the kinetic parameters for the parent compound could be described with greater certainty compared to fitting without metabolite data or compared to fitting with the two other error models available in mkin.

**Table 4.** Degradation model parameters obtained for coupled fits to Soil Dataset 3.

Parameter	OLS		Variance by Variable		Two-Component	
	Estimate	Pr(>t)	Estimate	Pr(>t)	Estimate	Pr(>t)
parent_0	100	$5.3 \times 10^{-37}$	100	$2.9 \times 10^{-33}$	101	$5.04 \times 10^{-30}$
k1	0.0218	$8.65 \times 10^{-8}$	0.0261	0.00358	0.0294	0.00115
k2	$1.22 \times 10^{-5}$	0.499	0.00707	0.199	0.00826	0.0407
g	0.901	$2.8 \times 10^{-7}$	0.697	0.0477	0.615	0.0162
f_parent_to_A1	0.299	$3.16 \times 10^{-10}$	0.298	$2.45 \times 10^{-12}$	0.288	$2.2 \times 10^{-12}$
k_A1	0.0204	$2.98 \times 10^{-7}$	0.0204	$2.44e \times 10^{-9}$	0.0195	$1.01 \times 10^{-9}$

#### 4. Conclusions

The technical feasibility of using alternative error models for the kinetic evaluation of chemical degradation data was demonstrated. A two-component error model derived from experience gathered in analytical chemistry was proposed as a valid alternative to unweighted least squares fitting and to the variance by variable approach commonly used in iteratively reweighted least squares fitting for datasets including transformation products.

For reliable fitting of the components of the two-component error model while optimizing the degradation model parameters, it was found necessary to combine direct optimization of the likelihood function with a three-step procedure, consisting of an unweighted least squares fit to obtain degradation model parameters, a separate optimization of error model parameters with fixed degradation model parameters, and a final optimization of all parameters.

The AIC proved to be valuable for comparison of the error models, but also for comparison of the different degradation models and their combinations with the available error models.

For parent decline datasets and for datasets containing data for the parent and one or two metabolites, the two-component error model provided the most adequate representation of the data based on the AIC in one third of the cases or more.

In one case, a degradation model parameter was even significantly different from zero when using the two-component error model, while the parameter was not significantly different from zero under the other two error models.

In spite of the limited experience with the two-component error model, the results shown in this contribution suggested that it was a valid alternative to the error models used so far. However, it cannot be excluded that assessments will become more conservative if the method will be routinely applied.

**Supplementary Materials:** The following file is available online at <http://www.mdpi.com/2076-3298/6/12/124/s1>: PDF file showing the numerical results of the evaluations of the test datasets.

**Author Contributions:** Conceptualization, J.R. and S.M.; formal analysis, J.R.; methodology, J.R.; project administration, S.M.; software, J.R.; visualization, J.R.; writing, original draft, J.R.; writing, review and editing, S.M.

**Funding:** This research was funded by the German Environment Agency (Umweltbundesamt) Grant Number 112407.

**Acknowledgments:** The support by the German Environment Agency, both financially and by being interested in these matters, is gratefully acknowledged. We would like to thank Thomas Gräff for comments on an earlier version of the manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Liess, M.; Ratte, T.; Ebke, P.; Hollert, H. 20 years SETAC GLB: Increasing realism of pesticide risk assessment. *Environ. Sci. Eur.* **2019**. [[CrossRef](#)]
2. Knäbel, A.; Stehle, S.; Schäfer, R.B.; Schulz, R. Regulatory FOCUS Surface Water Models Fail to Predict Insecticide Concentrations in the Field. *Environ. Sci. Technol.* **2012**, *46*, 8397–8404. [[CrossRef](#)] [[PubMed](#)]
3. Knäbel, A.; Meyer, K.; Rapp, J.; Schulz, R. Fungicide Field Concentrations Exceed FOCUS Surface Water Predictions: Urgent Need of Model Improvement. *Environ. Sci. Technol.* **2014**, *48*, 455–463. [[CrossRef](#)] [[PubMed](#)]
4. Knäbel, A.; Scheringer, M.; Stehle, S.; Schulz, R. Aquatic Exposure Predictions of Insecticide Field Concentrations Using a Multimedia Mass-Balance Model. *Environ. Sci. Technol.* **2016**, *50*, 3721–3728. [[CrossRef](#)] [[PubMed](#)]
5. Boesten, J. Effects of Random and Systematic Errors on Freundlich Parameters for Pesticide Sorption. *Soil Sci. Soc. Am. J.* **2015**, *79*, 1306–1318. [[CrossRef](#)]
6. Gao, Z.; Green, J.; Vanderborght, J.; Schmitt, W. Improving uncertainty analysis in kinetic evaluations using iteratively reweighted least squares. *Environ. Sci. Technol.* **2011**, *45*, 4429–4437. [[CrossRef](#)] [[PubMed](#)]
7. Görlitz, L.; Gao, Z.; Schmitt, W. Statistical analysis of chemical transformation kinetics using Markov-Chain Monte Carlo Methods. *Environ. Sci. Technol.* **2011**, *45*, 4429–4437. [[CrossRef](#)] [[PubMed](#)]
8. Honti, M.; Fenner, K. Deriving Persistence Indicators from Regulatory Water-Sediment Studies—Opportunities and Limitations in OECD 308 Data. *Environ. Sci. Technol.* **2015**, *49*, 5879–5886. [[CrossRef](#)] [[PubMed](#)]
9. Honti, M.; Hahn, S.; Hennecke, D.; Junker, T.; Shrestha, P.; Fenner, K. Bridging across OECD 308 and 309 Data in Search of a Robust Biotransformation Indicator. *Environ. Sci. Technol.* **2016**, *50*, 6865–6872. [[CrossRef](#)] [[PubMed](#)]
10. Witt, J.; Gao, Z.; Meyer, H. KinGUI v2.1. 2014. Available online: <http://kingui.vrbka.net/KinGUIv2.1.zip> (accessed on 9 December 2019).
11. Tessella Technology and Consulting. CAKE 3.3: Computer Assisted Kinetic Evaluation. 2016. Available online: <https://www.tessella.com/showcase/computer-assisted-kinetic-evaluation> (accessed on 9 December 2019).
12. Werner, M.; Brooks, S.H.; Knott, L.B. Additive, Multiplicative, and Mixed Analytical Errors. *Clin. Chem.* **1978**, *24*, 1895–1898. [[PubMed](#)]
13. Schwartz, L.M. Calibration Curves with Nonuniform Variance. *Anal. Chem.* **1979**, *51*, 723–727. [[CrossRef](#)]
14. Rocke, D.M.; Lorenzato, S. A two-component model for measurement error in analytical chemistry. *Technometrics* **1995**, *37*, 176–184. [[CrossRef](#)]
15. Wilson, M.D.; Rocke, D.M.; Durbin, B.; Kahn, H.D. Detection limits and goodness-of-fit measures for the two-component model of chemical analytical error. *Anal. Chim. Acta* **2004**, *509*, 197–208. [[CrossRef](#)]
16. Ranke, J. mkin 0.9.49.7: Kinetic Evaluation of Chemical Degradation Data. Available online: <https://pkgdown.jrwb.de/mkin> (accessed on 9 December 2019).
17. FOCUS. Guidance Document on Estimating Persistence and Degradation Kinetics from Environmental Fate Studies on Pesticides in EU Registration. Report of the FOCUS Work Group on Degradation Kinetics, EC Doc. Ref. Sanco/10058/2005, version 2.0, Work Group on Degradation Kinetics of FOCUS (FORum for the Co-ordination of Pesticide Fate Models and Their USe). 2006. Available online: <https://esdac.jrc.ec.europa.eu/projects/degradation-kinetics> (accessed on 9 December 2019).
18. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
19. Davidian, M.; Carroll, R.J. Variance Function Estimation. *J. Am. Stat. Assoc.* **1987**, *82*, 1079–1091. [[CrossRef](#)]

20. FOCUS. Generic guidance for estimating persistence and degradation kinetics from environmental fate studies on pesticides in EU registration. Report Version 1.1, Work Group on Degradation Kinetics of FOCUS (FORum for the Co-Ordination of Pesticide Fate Models and Their USE). 2014. Available online: <https://esdac.jrc.ec.europa.eu/projects/degradation-kinetics> (accessed on 9 December 2019).
21. Ranke, J.; Wöltjen, J.; Meinecke, S. Comparison of software tools for kinetic evaluation of chemical degradation data. *Environ. Sci. Eur.* **2018**. [[CrossRef](#)] [[PubMed](#)]
22. Burnham, K.P.; Anderson, D.R. Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociol. Methods Res.* **2004**, *33*, 261–304. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).