

Kansas State University Libraries

New Prairie Press

---

Conference on Applied Statistics in Agriculture

1999 - 11th Annual Conference Proceedings

---

## MODELING THE FATE OF TOXIC CHEMICALS IN SOILS

Anabayan Kessavalou

Anne M. Parkhurst

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

---

### Recommended Citation

Kessavalou, Anabayan and Parkhurst, Anne M. (1999). "MODELING THE FATE OF TOXIC CHEMICALS IN SOILS," *Conference on Applied Statistics in Agriculture*. <https://doi.org/10.4148/2475-7772.1258>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact [cads@k-state.edu](mailto:cads@k-state.edu).

## MODELING THE FATE OF TOXIC CHEMICALS IN SOILS

Anabayan Kessavalou  
Monsanto

Anne M. Parkhurst  
Department of Biometry  
University of Nebraska-Lincoln

### ABSTRACT

Studies on the fate of toxic chemicals in soils are often reported with a minimum of descriptive statistics. Use of modeling techniques to describe the kinetics of chemical degradation provides a better understanding of the fate of chemicals in soil systems. When modeling nonlinear systems, assumptions made about the error term greatly influence the parameter estimation. Inappropriate use of linearization and failure to account for autocorrelated errors may result in inaccurate models. Information is also needed about the effects of the magnitude of autocorrelation on parameter estimation. The exponential decay function was chosen to fit the data obtained from a TNT (2, 4, 6-trinitrotoluene) degradation experiment in soil using four different error assumptions. Estimates of the rate constant ( $k$ ) and other parameter estimates changed appreciably as assumptions about the error term changed. Simulation studies indicated that modeling data from chemical decomposition studies with an independent error assumption resulted in unreliable  $k$  estimates when the autocorrelation was large. A two-step procedure was used to fit an exponential autocorrelated (AR(1)) model. Overall, the exponential function with the additive-correlated error assumption provided the best fit for TNT degradation data. In essence, the kinetic rate constant obtained through model fitting in chemical decomposition studies provides a great deal of useful information to scientists. However, the researcher must be aware of the fact that making correct assumptions about the error term is extremely critical for obtaining accurate and precise estimates of  $k$ .

### 1. INTRODUCTION

Improper use of chemicals in agricultural and industrial sectors often leads to the contamination of soil and water with toxic chemicals that are harmful to human and animal health (Alexander, 1999). Therefore, a thorough understanding of their behavior in soil is essential to develop techniques for remediating contaminated sites. To achieve this, soil chemists and environmental toxicologists routinely conduct laboratory and field experiments researching the fate of toxic chemicals in soils. In such studies, a known concentration of a chemical is added to the soil and effects of selected treatments on degradation kinetics are determined with time. Rate constants are then used to predict cleanup times for remediation techniques.

## 2. STATISTICAL ANALYSIS AND MODEL FITTING

Data obtained from such studies are often analyzed with graphical and descriptive statistics. But determining an apparent kinetic rate constant through model fitting gives an estimate of the rate of decomposition of the chemical, besides providing a better understanding of its behavior in soil (Alexander, 1999). In addition, apparent rate constants and information obtained from other experiments involving physical, chemical and biological interactions of the chemical in soil may be integrated to develop large-scale process models or decision support systems. Such tools can be used to tailor site-specific and efficient management practices for remediating contaminated soils.

The rate constant is estimated through model fitting. Making appropriate assumptions about the error term of the selected model is one of the key steps necessary for obtaining reliable parameter estimates. Usually the error term in the model selected is assumed additive, independent, identical and normal. But these assumptions may not hold for all situations, as error could be multiplicative, correlated, heterogeneous and non-normal depending upon the nature of the study and response variable measured. Failing to make appropriate assumptions would result in biased estimates of the parameters and standard errors. As a result, management decisions made based on such estimates will be poor and unreliable. The objective of this paper is to demonstrate the impact of ignoring and / or violating appropriate assumptions about the error term on the estimate of the rate constant using a two-parameter exponential model to fit data obtained from a TNT (2, 3, 6 trinitro toluene) degradation experiment and a simulation study.

## 3. METHODOLOGY

### TNT Experiment

TNT, an explosive chemical, was used to make bombs during World War II and the Korean War. Improper disposal of the TNT waste during the manufacturing of munitions led to the contamination of soil and ground water in several locations in the United States. An attempt to remediate these contaminated sites found  $\text{Fe}^0$  (zero-valent iron) was capable of abiotically degrading TNT in contaminated water and soil (Hundal et al., 1997). A laboratory study was conducted by Hundal et al. (1997) with different concentrations of  $\text{Fe}^0$ , (0.1, 0.2, 1, 2 and 10 % w/v), to determine the best  $\text{Fe}^0$  concentration to effectively degrade TNT in solution. The solution was treated with a known concentration of  $^{14}\text{C}$  labeled TNT and the decline in  $^{14}\text{C}$  concentration was monitored at 0, 0.5, 1, 2, 4, 8, 24, 48, 72 and 96 hours following the addition of  $\text{Fe}^0$ .

**Model Fitting and Design of the Study**

Graphical observation of the data revealed an exponential decay of <sup>14</sup>C concentration with time; the trend becoming prominent with increasing concentrations of Fe<sup>0</sup> (Fig. 1). Therefore, an exponential decay model of the functional form given below was chosen to fit the data, where the parameter of interest is k as it describes the rate of decomposition of TNT.

$$y = \beta \exp(-kt)$$

In the above equation, y is the relative concentration of TNT, β is the relative concentration of TNT at time 0, and k is the apparent rate of TNT degradation. The value of k can be estimated in two ways. The first approach is linearizing the model by taking the natural logarithm and estimating k through simple linear regression. Alternatively, k can be estimated through non-linear regression techniques without linearizing the model. Although both these approaches are mathematically equivalent, statistically they are not comparable. They differ markedly by their error assumptions. Linearization implies that the error is multiplicative, while estimating k using the nonlinear model assumes the error is additive. Statistically these models result in different estimates of k and its standard error.

Further, data in this study were collected from the same experimental unit over time and therefore it is likely that the errors are correlated. Ignoring this assumption can affect parameter estimation. Some of the consequences of assuming independent error when it is correlated are severe underestimation of standard error of k and the estimated confidence interval, and lack of precision in hypothesis testing. Therefore, four different exponential models, each with different error assumptions were fit to the TNT data to evaluate the impact of various error assumptions on estimates of k and its standard error. Also, a simulation study was conducted to determine the effects of k-values and the size of autocorrelation on estimates of k, as influenced by different error assumptions.

**Model 1.**

Model 1 was an additive and independent error model (Eq. 1). Rate constant, k was estimated using Proc NLIN with Marquardt’s method (SAS Institute, 1996). Initial values for β and k were obtained through linear regression techniques.

$$y_t = \beta \exp(-kt) + \epsilon_t \tag{Eq. 1}$$

$$\epsilon_t \sim \text{iidN}(0, \sigma^2)$$

**Model 2**

Model 2 was a multiplicative and independent error model that is given below.

$$y_t = \beta \exp(-kt) \epsilon_t .$$

Rate constant, k, was estimated by fitting the linearized exponential function shown below (Eq. 2) using Proc REG of SAS (SAS Institute, 1996).

$$\ln(y_t) = \ln(\beta) - kt + \ln \epsilon_t \tag{Eq. 2}$$

$$\ln \epsilon_t \sim \text{iidN}(0, \sigma^2)$$

Model 3

Model 3 was an additive and correlated error model. First order autocorrelation (AR (1)) error structure was assumed to model the data. Although time was unequally spaced no heterogeneous first-order autoregressive covariance (ARH(1)) was assumed, as more degrees of freedom will be lost in estimating unknown variance components when compared to AR (1) covariance structure. This could lead to inefficient estimation of standard errors and hypothesis testing, as the data set contained only 10 observations. In order to eliminate any such problems, AR (1) covariance was used to model the data. A general model structure is shown as follows:

$$y_t = \beta \exp(-kt) + a_t$$

$a_t$  (autocorrelated error) =  $\rho a_{t-1} + \epsilon_t$   
 $\rho$  = size of autocorrelation  
 $\epsilon_t$  = error at time t

The final exponential model with additive and correlated error was obtained as suggested by Bates and Watts (1988).

$$y_t = \beta * \exp(-kt) + \rho(y_{t-1} - \beta \exp(-kt_{t-1})) + \epsilon_t \tag{Eq. 3}$$

$$\epsilon_t \sim \text{iidN}(0, \sigma^2)$$

For this model, parameter estimates were obtained by using a two-step procedure following a procedure demonstrated in Gallant (1987). First, initial values for k and  $\beta$  were obtained by fitting an additive and independent error model (Eq. 1) to the data using Proc NLIN of SAS. Initial estimate for  $\rho$  was obtained by fitting AR(1) model using Proc ARIMA of SAS to the residuals obtained from the non-linear regression analysis. The second step involved obtaining final estimates for all three parameters (k,  $\beta$ , and  $\rho$ ) by fitting the additive and correlated model (Eq. 3) using Proc NLIN.

Model 4

Model 4 was an exponential function with multiplicative and AR(1) correlated error (Eq. 4). Parameter estimates were obtained by using Proc Autoreg of SAS.

$$\ln(y_t) = \ln(\beta) - kt + \ln(a_t) \tag{Eq. 4}$$

$a_t$  (autocorrelated error) =  $\rho a_{t-1} + \epsilon_t$   
 $\rho$  = size of autocorrelation  
 $\epsilon_t$  = error at time t

**Simulation study**

Data sets with different values of  $k = \{0.6, 0.4, 0.2, 0.1\}$  and an autocorrelation of  $\rho = 0.9$  were simulated as the exponential decay function with additive correlated error using SAS IML procedures. Each data set contained 1000 realizations of 10 observations per realization for each  $\rho$  and k combination. The error, assumed to be normal and additive, was generated using RANNOR and the random seed option in SAS. All four models described above were fit to the data and the rate constant k was estimated.

Sampling properties of the estimate of  $k$  were evaluated using the average value of  $k$ , standard error, bias of  $k$  (%), MSE, percent coverage by 95% asymptotic confidence intervals and the coefficient of skewness.

#### 4. RESULTS AND DISCUSSION

##### **Impact of modeling assumptions on $k$ : TNT data**

The kinetic rate constant,  $k$ , is a measure of apparent rate of TNT degradation in solution as influenced by  $\text{Fe}^0$  concentration. The apparent rate constant obtained through model fitting does not only provide quantifiable and more useful information on TNT degradation when compared to other analytical approaches such as graphical and descriptive statistics, but also facilitates treatment comparisons. When the error was assumed independent, estimates of  $k$  obtained by fitting the additive error model (model 1) ranged from 0.008 to 0.303, while the multiplicative error model (model 2) yielded  $k$  ranging from 0.008 to 0.041 (Table 1). When correlated error was assumed, the additive error model (model 3) gave  $k$  values comparable to model 1 and the multiplicative model (model 4) gave  $k$  values comparable to model 2 (Table 1).

Regardless of error assumptions, absolute values of  $k$  increased with increasing concentrations of  $\text{Fe}^0$ . However, magnitude of  $k$  values estimated with model 1 (additive and independent) and 3 (additive and correlated) were as high as 13 times greater than that estimated with model 2 (multiplicative and independent) and model 4 (multiplicative and correlated). Further, model mean square errors (MSE), used as a tentative measure of "fit", were several folds smaller for additive than for multiplicative model, irrespective of independent and correlated error structures (Table 1). Examination of residuals, non-zero parameter estimates, and smaller MSE suggest that the exponential decay functions with additive error term provided better models for this data set.

Because the models with additive error assumption appeared to better describe the data than the multiplicative models, parameter estimates of independent and correlated error model with additive error were compared. Slightly larger estimates of rate constant ( $k$ ) and lower MSE were obtained with correlated error model (model 3) than with the independent error model (model 1) (Table 2). Further, no violation of assumptions was detected in the distribution of residuals from the correlated error model. Standard error of  $k$  estimated with the correlated error model ranged from 0.039 to 0.098 and it was twice as big as that of the independent error model. Smaller standard error values for  $k$  in the independent error model may be attributed to the underestimation of standard error possibly because of the failure to account for correlation among errors (Seber and Wild, 1989). The estimates of autocorrelation ranged from 0.82 to 0.99. Further, the upper and lower limits of the 95% asymptotic confidence interval for  $k$  obtained with independent error model (model 1) did not encompass the values of  $k$  estimated with autocorrelated error model (Table 2). Thus, it is possible that estimates of  $k$  obtained using a false independent error assumption may deviate significantly from its true value resulting in misleading and unrealistic conclusions about the rate of TNT destruction by  $\text{Fe}^0$ . Serious

practical consequences may arise if such unreliable information was used to develop remediation practices.

Results of the TNT data analysis clearly demonstrate that assumptions made about the error term play a significant role in model fitting. As shown here, estimates of model parameters could change appreciably depending upon the error assumptions. Therefore, it is important to make appropriate assumptions about the error term to get correct parameter estimates. Overall, results indicate that this data set could be modeled appropriately using exponential decay function with additive and correlated error structure (model 3).

### **Impact of modeling assumptions on k:Simulation study**

The performance of the four models was compared using data simulated with an autocorrelation of 0.9 and  $k$  values of 0.6, 0.4, 0.2, and 0.1 (Table 3). Regardless of  $k$  values, additive models performed better than the multiplicative models. In the case of correlated model,  $k$  was estimated with a bias ranging from 4 to -7% using additive error model compared to the bias of -60 to -93% with multiplicative model. Lack of skewness in the distribution of  $k$ , low MSE and no violation of assumptions in the distribution of residuals (not presented) also strongly supported the use of the exponential function with additive error model.

As the data were generated with high autocorrelation (0.9), comparison between the independent and correlated model (Table 3) is of primary interest. Rate constants ( $k$ ) were estimated more accurately and precisely using correlated error model than independent error, regardless of the size of  $k$ . For instance, percent bias in  $k$ , MSE and skewness were -7%, 0.001, and 0.53, respectively, for correlated error model when compared to -35%, 0.007, and 0.419, respectively, for independent error model when the size of  $k$  was 0.6. The percentage of 95% asymptotic confidence intervals that covered the true value of  $k$  is shown in Table 4. For the correlated model with  $k=0.6$ , 84% of the confidence intervals covered  $k$  while the coverage was only 25% in the case of independent error model. Failure to obtain close to 95% coverage for the correlated model may be attributed to the sample size of 10 observations, as the asymptotic confidence interval is computed based approximations to linear behavior. The impact of this limitation became severe when incorrect assumptions were made about the error term.

As  $k$  values can change with treatments, the impact of various error assumptions on  $k$  estimation was evaluated for different values of  $k$ . At high autocorrelation (0.9), regardless of size of  $k$ , bias in  $k$  remained low and stable when  $k$  was estimated with correlated-additive error model (Table 3 and Fig. 2). The increase in percent bias with increasing values of  $k$  in the case of independent-additive model could be attributed to severe violation of assumption of linearity in parameter space of  $k$  possibly made worse by correlated errors. Changes in size of  $k$  did not affect the percent coverage of  $k$  values in 95% asymptotic confidence interval when  $k$  was estimated using correlated-additive error model (Table 4). On the other hand, with independent-additive error model, percent coverage decreased dramatically as  $k$  increased.

More simulations are necessary to evaluate the effects of the size of autocorrelation on parameter estimates. However, for a strong autocorrelation ( $\rho=0.9$ ), results indicate that the

two-step procedure described for fitting model 3 provides an adequate method for fitting an exponential autocorrelated (AR(1)) model.

## **5. CONCLUSIONS**

In chemical decomposition studies, estimating the kinetic rate constant using model-fitting techniques is very informative and beneficial. In doing so, it is imperative that scientists be aware of the importance of making appropriate assumptions regarding the error term in the model in order to accurately and precisely estimate  $k$ , as  $k$  values and other inferential properties change with error assumptions. Failure to use correct error assumptions could lead to misleading results as outlined in this study. It appears that the two-step procedure provides a means for fitting exponential autocorrelated (AR(1)) models to obtain reliable estimates of  $k$  irrespective of the size of  $k$  provided the autocorrelation is large.

## **ACKNOWLEDGEMENT**

Authors thank Drs. Steven D. Comfort and Lukwinder Hundal for providing the data set and Dr. Daryl Travnicek for helping with the SAS code

## **REFERENCES**

- Alexander, M., 1999. Kinetics. p. 73-101. *In* Biodegradation and Bioremediation. (ed). M. Alexander. Second edition. Academic Press. New York.
- Bates, M. B., and D.G. Watts, 1988. Nonlinear regression analysis and its applications. John Wiley & Sons, Inc., New York. Practical considerations in nonlinear regression. p.92-96.
- Gallant, R. 1987. Nonlinear statistical models. John Wiley & Sons., New York.
- Hundal, L., J. Singh, E. Bier, P.J. Shea, S.D. Comfort, and W.L. Powers. 1997. Removal of TNT and RDX from soil and water using iron metal. *Environ. Poll.* 97:55-64.
- SAS Institute, 1996. SAS users manual Version 6. Raleigh, North Carolina.
- Seber, G.A.F, and C.J. Wild, 1989. Nonlinear regression. John Wiley & Sons Inc., New York. Autocorrelated errors. p.271-322.



Table 1. Estimated rate constant (k) and mean square error (MSE) for the TNT data modeled with exponential function assuming various error assumptions.

Fe <sup>0</sup> conc., %	Additive error		Multiplicative error	
	k	MSE	k	MSE
<u>Independent error</u>				
	Model 1		Model 2	
0.1	0.008	0.001	0.008	0.002
0.2	0.021	0.006	0.017	0.014
1.0	0.115	0.005	0.034	0.139
2.0	0.132	0.006	0.032	0.148
10.0	0.303	0.004	0.041	0.564
<u>Correlated error</u>				
	Model 3		Model 4	
0.1	0.015	0.0003	0.007	0.008
0.2	0.035	0.002	0.015	0.006
1.0	0.170	0.003	0.029	0.069
2.0	0.199	0.002	0.025	0.067
10.0	0.413	0.002	0.030	0.563

Table 2. Parameter estimates obtained for TNT data by fitting the exponential function assuming independent and correlated additive error structure.

Fe <sup>0</sup> conc., %	MSE	k	Standard error	95% asymptotic confidence interval	
				Lower limit	Upper limit
<u>Additive independent error model (Model 1)</u>					
0.1	0.001	0.008	0.0006	0.007	0.009
0.2	0.006	0.021	0.003	0.014	0.027
1.0	0.005	0.115	0.018	0.073	0.156
2.0	0.006	0.132	0.024	0.077	0.186
10.0	0.004	0.303	0.042	0.205	0.401
<u>Additive correlated error model (Model 3)</u>					
0.1	0.0003	0.015	0.004	0.005	0.026
0.2	0.002	0.035	0.012	0.006	0.064
1.0	0.003	0.170	0.039	0.078	0.262
2.0	0.002	0.199	0.043	0.098	0.299
10.0	0.002	0.413	0.098	0.181	0.645

Table 3. Simulation results of all the four models for  $\rho=0.9$  and  $k=0.6, 0.4, 0.2$  and  $0.1$ .

Parameters	Correlated error		Independent error	
	Additive	Multiplicative	Additive	Multiplicative
<b><u>k = 0.6</u></b>				
Mean k	0.56	0.042	0.39	0.051
Standard error of k	0.082	0.012	0.061	0.010
Bias, %	-7	107	-35	92
Mean square error	0.001	0.373	0.007	0.579
Skewness	0.053	5.72	0.419	5.85
<b><u>k = 0.4</u></b>				
Mean k	0.38	0.040	0.29	0.049
Standard error of k	0.055	0.011	0.042	0.009
Bias, %	-5	-90	-28	-88
Mean square error	0.001	0.352	0.006	0.522
Skewness	0.323	5.61	0.465	5.60
<b><u>k = 0.2</u></b>				
Mean k	0.202	0.041	0.165	0.048
Standard error of k	0.028	0.008	0.023	0.006
Bias, %	1	-80	-18	-77
Mean square error	0.001	0.250	0.005	0.006
Skewness	0.591	4.26	0.692	5.34
<b><u>k = 0.1</u></b>				
Mean k	0.104	0.04	0.09	0.043
Standard error of k	0.017	0.006	0.011	0.005
Bias, %	4	-60	-10	-57
Mean square error	0.001	0.188	0.004	0.226
Skewness	0.403	2.61	0.525	2.72

Table 4. Percentage of 95 % asymptotic confidence intervals covering k for additive error model with correlated and independent error structure for  $\rho=0.9$

k-value	% coverage in 95% asy. CI	
	<u>Correlated</u>	<u>Independent</u>
	<u>error</u>	<u>error</u>
0.1	84	52
0.2	87	57
0.4	85	38
0.6	84	25

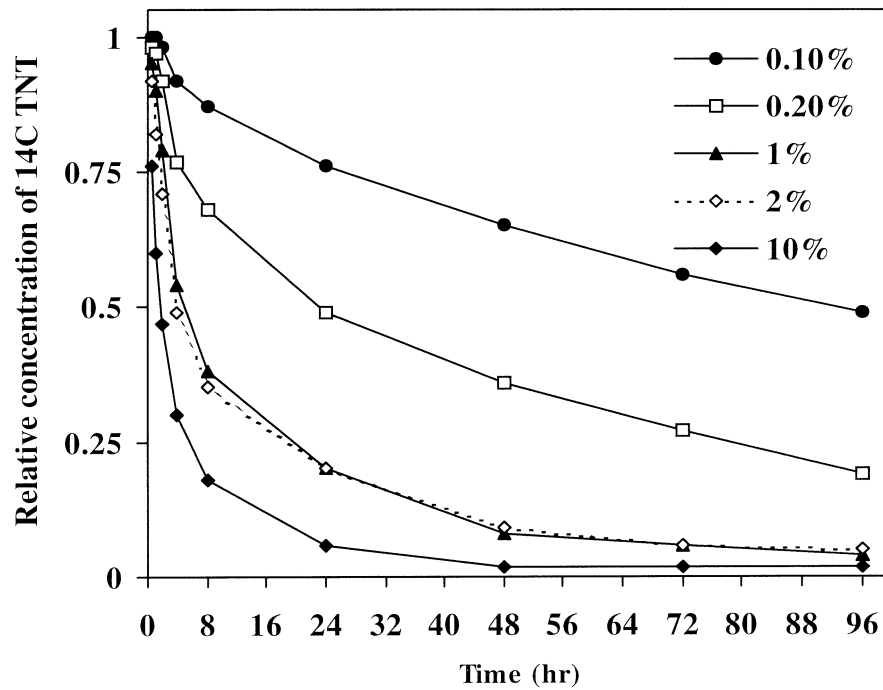
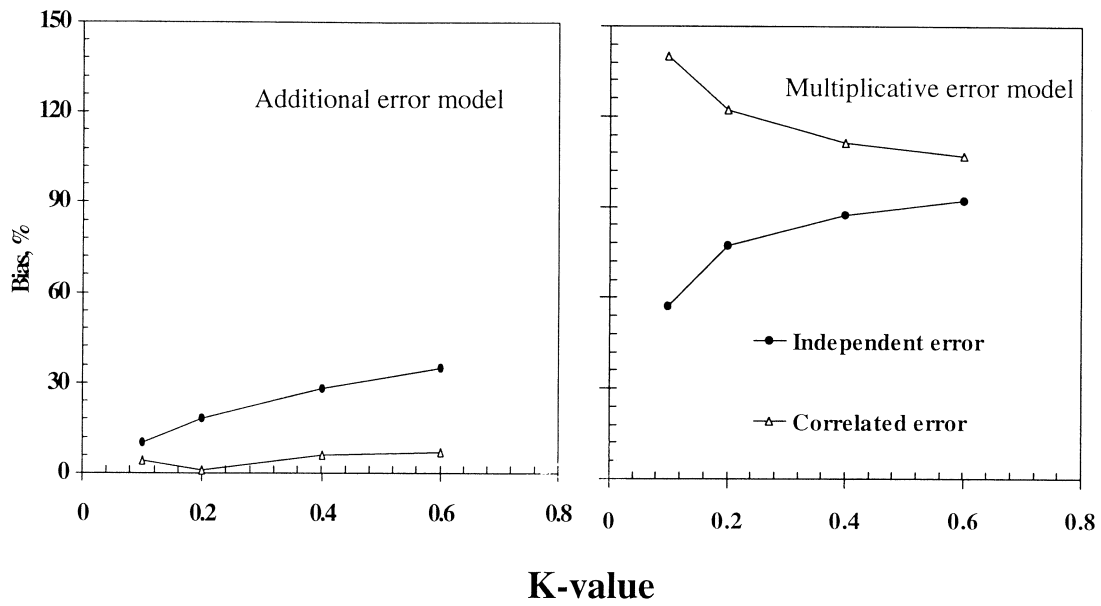


Figure 1. Loss of <sup>14</sup>C TNT as influenced by Fe<sup>0</sup> concentration (% w/v) at 25 °C



**Figure2. Bias as a function of K values**