

**CHEMICAL
RESEARCH,
DEVELOPMENT &
ENGINEERING
CENTER**

CRDEC-TR-88157

**INHERENT BIAS
IN NONLINEAR LEAST SQUARES ANALYSIS
OF FIRST-ORDER KINETIC DATA**

AD-A210 677

Barry L. Knier, Ph.D.
UNIVERSITY OF PUERTO RICO, RIO PIEDRAS CAMPUS
Rio Piedras, PR 00931

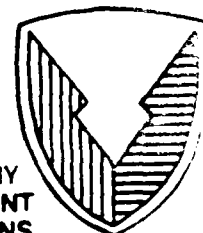
DTIC
ELECTE
JUL 19 1989
S D

Frederick R. Longo, Ph.D.
DREXEL UNIVERSITY
Philadelphia, PA 19104

June 1989

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

**U.S. ARMY
ARMAMENT
MUNITIONS
CHEMICAL COMMAND**



Disclaimer

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorizing documents.

Distribution Statement

Approved for public release; distribution is unlimited.

REPORT DOCUMENTATION PAGE

| | | | | | |
|--|--|-----------|--|----------------------------------|-------------------|
| 1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED | | | 1b RESTRICTIVE MARKINGS | | |
| 2a SECURITY CLASSIFICATION AUTHORITY | | | 3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited. | | |
| 2b DECLASSIFICATION/DOWNGRADING SCHEDULE | | | 5. MONITORING ORGANIZATION REPORT NUMBER(S) | | |
| 4 PERFORMING ORGANIZATION REPORT NUMBER(S) CRDEC-TR-88157 | | | 5. MONITORING ORGANIZATION REPORT NUMBER(S) | | |
| 6a NAME OF PERFORMING ORGANIZATION CRDEC | 6b OFFICE SYMBOL (If applicable) SMCCR-RSC-C | | 7a NAME OF MONITORING ORGANIZATION | | |
| 6c ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21010-5423 | | | 7b ADDRESS (City, State, and ZIP Code) | | |
| 8a NAME OF FUNDING/SPONSORING ORGANIZATION CRDEC | 8b OFFICE SYMBOL (If applicable) SMCCR-RSC-C | | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER | | |
| 8c ADDRESS (City, State, and ZIP Code) Aberdeen Proving Ground, MD 21010-5423 | | | 10 SOURCE OF FUNDING NUMBERS | | |
| | | | PROGRAM ELEMENT NO. | PROJECT NO. 1L162706 | TASK NO. A553F |
| | | | WORK UNIT ACCESSION NO. | | |
| 11 TITLE (Include Security Classification) Inherent Bias in Nonlinear Least Squares Analysis of First-Order Kinetic Data | | | | | |
| 12 PERSONAL AUTHOR(S) Knier, Barry L., Ph.D. (University of Puerto Rico), and (continued on reverse) | | | | | |
| 13a. TYPE OF REPORT Technical | 13b. TIME COVERED FROM 85 Jun to 86 Sep | | 14. DATE OF REPORT (Year, Month, Day) 1989 June | 15 PAGE COUNT 11 | |
| 16 SUPPLEMENTARY NOTATION | | | | | |
| 17 COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) | | |
| FIELD | GROUP | SUB-GROUP | First-Order Kinetics | | |
| 15 | 02 | | Decontamination, Error Analysis, | | |
| 07 | 03 | | Least Squares Analysis, Catalysis | | |
| 19. ABSTRACT (Continue on reverse if necessary and identify by block number) | | | | | |
| <p>Generally, it is assumed in analyzing first-order kinetic data that the unweighted fit to the exponential function, $b[\exp(-kt)]$, is superior to that of the logarithmic function, $\ln b - kt$. A careful analysis shows that this conclusion is not justified. In addition, a comparison of fits of actual data to both functions demonstrates that the methods are essentially equivalent.</p> | | | | | |
| 20 DISTRIBUTION/AVAILABILITY OF ABSTRACT <input checked="" type="checkbox"/> UNCLASSIFIED/UNLIMITED <input type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS | | | 21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED | | |
| 22a NAME OF RESPONSIBLE INDIVIDUAL SANDRA J. JOHNSON | | | 22b TELEPHONE (include Area Code) (301) 671-2914 | 22c OFFICE SYMBOL SMCCR-SPS-T | |

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

11. Personal Letter (continued)

John J. McDevitt, Jr., Ph.D. (Drexel University)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

PREFACE

The work described in this report was authorized under Project No. 1L162706A553F, CB Decontamination and Contamination Avoidance. This work was started in June 1985 and completed in September 1986.

The use of trade names or manufacturers' names in this report does not constitute an official endorsement of any commercial product. This report may not be cited for purposes of advertisement.

Reproduction of this document in whole or in part is prohibited except with permission of the Commander, U.S. Army Chemical Research, Development and Engineering Center, ATTN: SMCCR-SPS-T, Aberdeen Proving Ground, Maryland 21010-5423. However, the Defense Technical Information Center and the National Technical Information Service are authorized to reproduce the document for U.S. Government purposes.

This report has been approved for release to the public.



| | |
|---------------|---|
| Accession | |
| NTIS | ✓ |
| DTIC TAB | ✓ |
| Unannounced | ✓ |
| Justification | |
| by | |
| DTIC TAB | |
| Unannounced | |
| Justification | |
| DTIC | |
| Unannounced | |
| Justification | |
| A-1 | |

Blank

CONTENTS

| | Page |
|--|------|
| 1. INTRODUCTION..... | 7 |
| 2. UNCERTAINTY ANALYSIS FOR LINEAR FUNCTIONS..... | 8 |
| 3. UNCERTAINTY ANALYSIS FOR NONLINEAR FUNCTIONS..... | 8 |
| 4. CONCLUSION..... | 10 |
| LITERATURE CITED..... | 11 |

Blank

INHERENT BIAS IN NONLINEAR LEAST SQUARES ANALYSIS OF FIRST-ORDER KINETIC DATA

1. INTRODUCTION

It has become increasingly popular to use an unweighted nonlinear least squares analysis to evaluate first-order rate constants.¹⁻³ The typical procedure starts with a data set t_i, y_i and fits it to a function

$$p_t = a + be^{-kt} \quad (1)$$

where p_i is the predicted (calculated) value of the measurement y_i . This is a relatively complex problem because the nonlinear fitting routine does not reduce algebraically (as does the linear version) so that various values of the constants a , b , and k are tried, and a minimum variance is sought. It is used in preference to its linear analog

$$\ln(y_i - a) = -kt + \ln(b) \quad (2)$$

because the experimental uncertainties of the y_i values do not transform linearly into \ln space. The uncertainty of $\ln(y_i - a)$ increases with time and in the unweighted procedure introduces a large potential error into the determination of k .

The shortcomings of the linear procedure are well known^{1,4-6} but no one has carefully examined the difficulties inherent in the exponential fit routine. There is the tacit assumption that, except for the mathematical gymnastics, the unweighted exponential fit routine is the perfect solution to the data analysis. Actually, it has the same difficulties as the linear case; they are just more deeply hidden.

The unweighted least squares procedure makes use of several assumptions. It assumes that all of the errors can be associated with one variable (y) and that errors are random (i.e., Gaussian). Goodness of fit is then determined by minimizing the sum of $(p_i - y_i)^2$, where p_i is the theoretical (calculated) value determined by substituting t_i into a trial

function [i.e., $p_i = f(t_i)$]. Practically a data set t_i, y_i is observed and matched to a calculated data set t_i, p_i .

Each term in the least squares sum

$$(y_i - p_i)^2 \quad (3)$$

is considered to be equally valid because each y_i is equally valid (i.e., has the same uncertainty). The problem arises from the first assumption: all the errors can be lumped into y . This implies that p_i is exact because it is calculated from t_i that is assumed to be exact. Because both t and y are experimental quantities, the validity of that assumption needs to be examined.

2. UNCERTAINTY ANALYSIS FOR LINEAR FUNCTIONS

For linear functions, there is no problem; and the assumption holds, viz.,

$$p_i = Ct_i \quad (4)$$

If each t_i is associated with an uncertainty u_t that is constant for all t_i , then the uncertainty of p_i (u_p) is given by

$$u_p = Cu_t \quad (5)$$

and is constant for all of the calculated p_i values; and each term of the least squares sum is equally precise.

3. UNCERTAINTY ANALYSIS FOR NONLINEAR FUNCTIONS

For nonlinear functions, the situation is more complex. For the exponential case, equation 3 becomes

$$p_i = b \exp(-kt_i) \quad (6)$$

If the uncertainty of each t_i (u_t) is constant, then the largest value that the calculated p_i can assume is

$$p_i \text{ (largest)} = b \exp[-k(t_i - u_t)] \quad (7)$$

$$= b \exp(-kt_i) \exp(ku_t)$$

and the smallest is

$$p_i \text{ (smallest)} = b \exp[-k(t_i + u_t)] \quad (8)$$

$$= b \exp(-kt_i) \exp(-ku_t)$$

Considering the error associated with p_i to be half the length of its error bar, then

$$2u_p = p_i \text{ (largest)} - p_i \text{ (smallest)} \quad (9)$$

$$= b \exp(-kt_i) \exp(ku_t) - b \exp(-kt_i) \exp(-ku_t)$$

$$= b \exp(-kt_i) [\exp(ku_t) - \exp(-ku_t)]$$

Since b , k , and u_t are constants, the bracketed factor is also a constant. Using A to represent collected constants, the net result is

$$u_p = A b \exp(-kt_i) \quad (10)$$

that, when compared to equation 6, becomes

$$u_p = A p_i \quad (11)$$

So rather than being a constant as is implicit in the normal application of the least squares analysis, u_p is a linear function of p_i or its experimental counterpart y_i .

Using relative values to relate to the chemical problem of first-order kinetics, y_i varies from 100-0%. Thus, points near 100% have twice the error as points near 50% and 10 times the error as points near 10%. Because the validity (i.e., appropriate weight) of a point relates to the inverse square of its uncertainty and because chemical reactions are typically followed for five half-lives (i.e., from 100-3%), this represents a serious problem for any type of unweighted analysis.

4. CONCLUSION

The problem relates to the change in slope (dy/dt) from a high initial value to zero at the completion of the reaction. During the first stages of the reaction, the concentration of the reactant, y , changes very rapidly. Any uncertainty of the time of a data point t in that region does not add any uncertainty to the y value of the point, but it does have a large effect on the corresponding p value because it is calculated directly from the time t . At later stages of the reaction where y is changing less rapidly, uncertainties in the time values add much less uncertainty to the calculated value of p . The net result is that the uncertainty of the t, y point is not constant for the duration of the reaction. No data analysis can be considered completely appropriate unless this is taken into account.

LITERATURE CITED

1. Swain, C.G., Swain, M.S., and Berg, L.F., "Calculation of Unbiased First-Order Rate Constants, Acceleration by Overrelaxation with a Responsive Overrelaxation Factor," J. Chem. Inf. Comput. Sci. Vol. 20, p 47 (1980).
2. DeTar, D.F., "Significant Figures and Correlation of Parameters," J. Chem. Ed. Vol. 44, p 759 (1967).
3. DeTar, D.F., Computer Programs for Chemistry Vol. 1, pp 117-125 and 126-173, W.A. Benjamin Publishing Company, New York, 1968.
4. Roseveare, W.E., "Methods of Calculating and Averaging Rate Constants," J. Amer. Chem. Soc. Vol. 53, p 1631 (1931).
5. Collins, C.J., "Isotopes and Organic Reaction Mechanisms," Adv. Phys. Org. Chem. Vol. 2, p 62 (1964).
6. Wold, S., "A Note on the Computation of Rate Constants by Steepest Descent and Related Methods," Acta Chem. Scand. Vol. 21, p 1986 (1967).