CORRELATION AND OVERCORRELATION OF HETEROGENEOUS REACTION RATE DATA

MORDECHAI SHACHAM, NEIMA BRAUNER*
Ben-Gurion University of the Negev • Beer-Sheva, 84105, Israel

Experimental data is hard to come by, and obtaining it can often be expensive, difficult, and time-consuming. But with the computational tools available today, it is easy to manipulate available data in order to extract every bit of information that exists. There are, however, dangers in manipulating the data. If it is not done carefully, the data can be altered without warning and unnoted, and it will no longer provide the right information. Another danger is overcorrelation, which is done by forcing a model when there is insufficient or inaccurate data, trying to get information that is not there.

How can we know what is the valid use of data and what represents unjustified altering or overcorrelation? The answer is: there is no need for complicated statistical analysis. Common sense and judicious use of some basic statistical and error-analysis tools provide the answer in most cases.

Most of today's students have access to programs which

Mordechai Shacham is Professor and Head of the Chemical Engineering Department at the Ben Gurion University of the Negev, Beer-Sheva, Israel. He received his BSc and DSc from the Technion, Israel Institute of Technology. His research interests include applied numerical methods, computer-aided instruction, chemical process simulation, design, and optimization, and expert systems.
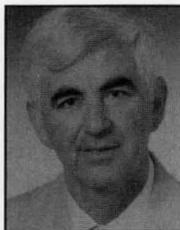
Neima Brauner received her BSc and MSc from the Technion, Israel Institute of Technology, and her PhD from the University of Tel-Aviv. She is currently Associate Professor in the Fluid Mechanics and Heat Transfer Department and has been elected to President of the Israel Institute of Chemical Engineers. She teaches courses in Mass and Heat Transfer and Process Control. Her main research interests include two-phase flows and transport phenomena in thin films.

* Address: School of Engineering, Tel-Aviv University, Tel-Aviv, 69978, Israel

carry out linear, polynomial, and often even nonlinear regression. While these programs fit the parameters to the requested model, they will not provide guidance regarding the ability of the model to represent the data. The critical analysis of the model's adequacy to represent the data must be done by the students, and we should provide them with the tools needed to carry out this analysis.

In this paper, we will use an example from the area of heterogeneous reaction rate data analysis to demonstrate some potential pitfalls in data correlation. Churchill noted that, "The published correlations for reactions catalyzed by solids provide many examples of overcorrelation."[1] The example we will be using involves verification of a rate expression for platinum catalyzed oxidation of sulfur dioxide. This example is from Smith's textbook[2] on chemical engineering kinetics.

SULFUR DIOXIDE OXIDATION EXAMPLE

The data of reaction rate versus partial pressures of SO_2, SO_3, and O_2 in oxidation of sulfur dioxide using a packed bed of platinum-on-alumina catalyst pellets at 480°C is shown in Table 1. The data is from Example 9.2 in Smith. The data shown in the table is actually part of a larger set of data which was published by Olson, *et al.*, in 1950.[3]

Based on a postulated reaction mechanism, Smith developed the following rate expression for this reaction.

$$r = \frac{p_{SO_2} p_{O_2}^{1/2} - (1/K)p_{SO_3}}{\left(A + Bp_{SO_3}\right)^2} \qquad (1)$$

where K is the equilibrium coefficient (K = 73 at 480°C) and A and B are constants to be calculated by regression of the experimental data. Smith calculated the constants using lin-

ear regression with the linearized form of Eq. (1)

$$\left[\frac{p_{SO_2} p_{O_2}^{1/2} - (1/K)p_{SO_3}}{r}\right]^{1/2} = A + Bp_{SO_3} \quad (2)$$

Since the unknown coefficients in Eq. (2) are in linear form, they can be calculated either by using linear regression or by plotting the expression on the left-hand side of this equation versus p_{SO_3} and fitting a straight line. The coefficient values obtained by Smith using this procedure are A = 0.176 and B = 12.9. The test for feasibility of a particular mechanism to represent the data, when a Langmuir-Hinshelwood type rate expression (such as Eq. 1) is used, is that all coefficients of the proposed model must be positive. In this case, the coefficients (A,B) are indeed positive; hence, the proposed rate expres-

sion is a feasible one.

We have recalculated the parameter values using nonlinear regression and the original Eq. 1. The calculation was done using the POLYMATH[4] package. The parameters obtained using nonlinear regression (including confidence intervals) are

$$A = 0.1017 \pm 0.0958$$

$$B = 16.02 \pm 4.33$$

These values are also positive. A comparison of the experimental and calculated reaction-rate values is shown in Figure 1.

It can be seen that the fit between the calculated and experimental values is very good. As an additional check, the sum of squares of errors, S^2, can be calculated

$$S^2 = \sum_{i=1}^{6} \left(r_{i,exp} - r_{i,calc}\right)^2$$

For the parameters obtained using nonlinear regression, $S^2 = 5.226 \times 10^{-5}$, a small number, indicating that indeed the fit between the experimental data and the calculated values is acceptable.

When all the indicators show that the fit is acceptable, does it ensure that Eq. (1) is the right rate expression to represent the reaction under the given conditions? Churchill warns that, "One should be wary of accepting the validity of a model merely because it is successful in correlating the data..." So what else should be checked?

Is the data really experimental data?

Looking at the data in Table 1 may have led one to suspect that this data does not represent the real measured values. The reaction rate, which is the dependent variable, appears in nice round numbers. While the value of the independent variable can often be set to rounded numbers, which are more convenient to work with, the measured (or calculated) value of the dependent variable will usually contain at least as many decimal digits as provided by the measuring instruments.

There is an explanation for the round numbers that represent the reaction rates in Table 1. Smith noted that the data was "interpolated for even intervals of reaction rates...." But when the original source of the data is consulted (Olson, et al.), one finds that this was not the only interpolation performed. Due to difficulties in controlling and measuring the temperature, none of the experiments was actually carried out at 480°C. The data in Table 1 is probably based on four measurements made at mean bed temperatures of 461°C, 467°C, 482°C, and 488°C. The results at these temperatures were again interpolated to 480°C. In 1950, interpolation meant putting the experimental data points on a graph paper and manually fitting the "best" curve. Hence, the sophisticated regression analysis is not carried out on the real measured data, but rather on a curve that was drawn freehand.

TABLE 1
Reaction Rate Data for
Sulfur Dioxide Oxidation at 480°C
(from Smith[2])

r g mol (h)(g catalyst)	Partial pressure (atm) at catalyst surface		
	SO_3	SO_2	O_2
0.02	0.0428	0.0255	0.186
0.04	0.0331	0.0352	0.190
0.06	0.0272	0.0409	0.193
0.08	0.0236	0.0443	0.195
0.10	0.0214	0.0464	0.196
0.12	0.0201	0.0476	0.197
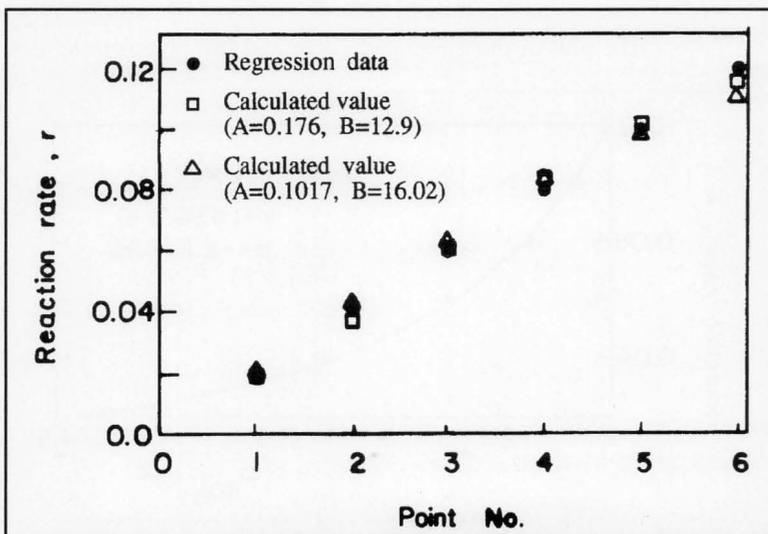
Figure 1. *Experimental and calculated reaction rate, using Eq. (1) as rate expression*

While we must realize that this was perhaps the best that could be done with the calculating tools that were available forty years ago, the approach is totally unacceptable today. Interpolation and extrapolation introduce changes in the error distribution of the data. Statistical analysis of regression models is based on certain assumptions regarding the error distribution. Therefore, with such smoothed and extrapolated data, the application of statistical analysis for evaluating the validity of regression models is meaningless.

Can the rate data be represented better by a different model?

Since the equilibrium coefficient value is $K = 73$, it seems that the reversibility of the reaction has very little significance at this temperature. Actually, it can be seen that the contribution of the reverse reaction (in the numerator of Eq. 1) is smaller by two orders of magnitude than the contribution of the forward reaction. This possibly suggests that the expression related to the reversible reaction can be omitted from the rate expression. Thus, we may try to find out the coefficients for the following rate expression:

$$r = \frac{p_{SO_2} p_{O_2}^{1/2}}{\left(A + B p_{SO_3}\right)^2} \tag{3}$$

The calculated coefficients using nonlinear regression are

$$A = 0.094 \pm 0.093$$
$$B = 16.53 \pm 4.20$$

with

$$S^2 = 4.8 \times 10^{-5}$$

The coefficients are positive and the sum of squares of errors is smaller than that obtained for Eq. (1). Thus, Eq. (3) is a valid rate expression and represents the data even better than Eq. (1). Is it the best rate expression?

Brauner and Shacham[5] have shown that the parameters of a power-law rate expression can be used to discriminate between feasible and infeasible mechanism-based rate expressions. For the reaction considered here, a power-law model reads

$$r = k p_{SO_3}^a \, p_{SO_2}^b \, p_{O_2}^c \tag{4}$$

By denoting α_{min} the lowest power of p_{SO_3} in the expression obtained for $1/r$ in a particular feasible mechanism-based model and α_{max} its highest power (according to the criteria developed by Brauner and Shacham[5]), the following inequality should hold:

$$\alpha_{min} < -a < \alpha_{max} \tag{5}$$

Similar criteria are valid for the powers of p_{SO_2} and p_{O_2}. The following results are obtained using nonlinear regression for the power-law model:

$$k = 0.517 \pm 113.3$$

$$a = -1.98 \pm 7.02$$
$$b = -0.216 \pm 4.556$$
$$c = 6.078 \pm 124.7$$
$$S^2 = 1.85 \times 10^{-5}$$

It can be seen that S^2 is considerably reduced compared to the values obtained in the two previously considered rate expressions, indicating that the power-law correlates the data much better than the previously suggested models.

We could proceed and postulate different feasible mechanism-based models using Eq. (5) and similar inequalities for the powers of the various partial pressures. But we should first pay attention to the huge uncertainty in the constants of the power-law model. The 95% confidence interval in k, for example, is larger by more than two orders of magnitude than the value of k itself. At this point, the statistical information on confidence limits becomes very important. Too much uncertainty in the parameter values may indicate that one or more of the variables should not be included in the correlation.

Since the uncertainty is the largest in the parameter c (the power of p_{O_2}), we can try first to remove p_{O_2} from the correlation by setting c = 0. The results obtained for this case are

$$k = 1.217 \times 10^{-5} \pm 1.268 \times 10^{-3}$$
$$a = -2.308 \pm 1.23$$
$$b = -0.052 \pm 1.87$$
$$S^2 = 1.89 \times 10^{-5}$$

The value of S^2 is changed very little by removing p_{O_2} from the correlation, but the confidence limits still remain too wide. This time the parameter b is associated with the largest uncertainty, so we may try now to also remove p_{SO_2} from the correlation by setting b = 0. The results obtained now are

$$k = (1.62 \pm 1.37) \times 10^{-5}$$
$$a = -2.270 \pm 0.222$$
$$S^2 = 1.9 \times 10^{-5}$$

The plot of the calculated curve and the experimental data
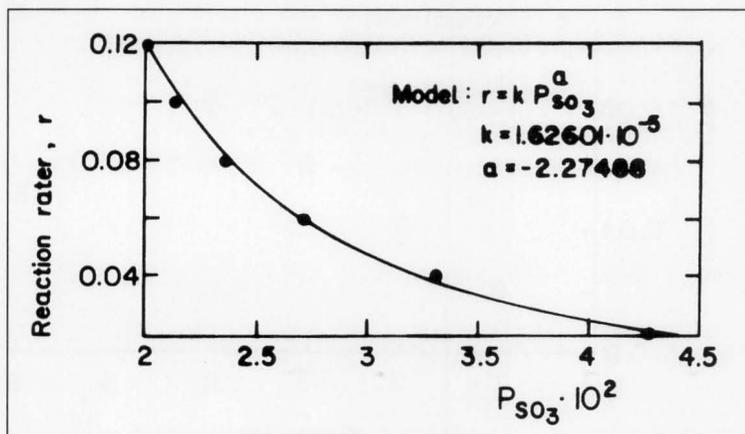
Figure 2. *Calculated curve and experimental data versus* p_{SO_3} *with power-law expression*

versus p_{SO_3} are shown in Figure 2.

All the indicators (plot, S^2, and confidence limits) show that so far this equation is the best to represent the data in Table 1. It should be emphasized, however, that this equation cannot be extrapolated outside the region where the measurements were made; otherwise, absurd rate values may result. For instance, at the start of the reaction when products' concentration is zero, this equation predicts an infinite reaction rate.

The poor asymptotic behavior of the power-law rate expression, obtained in this case, demonstrates very clearly the disadvantage of empirical models compared to mechanism-based models, which are derived based on physical insight. The power-law parameters, however, can be used for postulating mechanism-based rate expressions.

Inequality Eq. (5) with the power-law constants obtained above yields $\alpha_{min} < 2.27 < \alpha_{max}$. Using this inequality, several rate expression of the form of Eq. (3) can be postulated. One possibility is

$$r = \frac{1}{\left[A + Bp_{SO_3} \right]^3} \qquad (6)$$

Nonlinear regression with Eq. (6) yields

$$
\begin{aligned}
A &= 0.586 \pm 0.207 \\
B &= 72.49 \pm 9.35 \\
S^2 &= 2.56 \times 10^{-5}
\end{aligned}
$$

The plot of calculated curve and experimental data versus partial pressure of SO_3 for this correlation is practically identical to Figure 2, and thus this model also represents the data excellently.

The results for the different models are summarized in Table 2. They strongly suggest that the only variable to be included in the correlation of the data is p_{SO_3}. This is definitely true for the numbers appearing in Table 1, but will it hold in general for the sulfur dioxide oxidation reaction?

Why does the reaction rate depend only on p_{SO_3}?

Statistical analysis has shown that for the data of Table 1 the reaction rate does not depend on either p_{SO_2} or p_{O_2}. But before generalizing this conclusion, we have to seek the answer to two questions concerning experimental design:

1. Were these variables altered during the experiments to cause significant differences in the reaction rate?

2. Were the variables altered independently so that the data includes independent information with regard to their effect on the reaction rate?

Inspection of the partial pressures in Table 1 reveals that the relative variation of p_{O_2} is much smaller than that of p_{SO_2} and p_{SO_3}. The maximal change in p_{O_2} is less than 6%, while the changes in the other variables are of the order of 100%. But when considering the absolute change, it is the same order of magnitude for all the independent variables. Thus, the answer to the first question above is inconclusive.

To check the assumption that the partial pressures of the various components were altered independently, plots of one independent variable versus the others are prepared. The resulting correlation between p_{SO_3} and p_{SO_2} is shown in Figure 3.

TABLE 2
Summary of Fitted Parameter Values

Langmuir-Hinshelwood Type Model

Eq. #	A	B	S^2
1	0.102 ± 0.0958	16.02 ± 4.33	5.23×10^{-5}
3	0.094 ± 0.093	16.53 ± 4.2	4.8×10^{-5}
6	0.586 ± 0.207	72.49 ± 9.35	2.56×10^{-5}

Power-Law Model

k	a	b	c	S^2
0.517 ± 113.3	-1.98 ± 7.02	-0.216 ± 4.556	6.078 ± 124.7	1.85×10^{-5}
$(0.0121 \pm 1.27) \times 10^{-3}$	-2.308 ± 1.23	-0.052 ± 1.87	0	1.89×10^{-5}
$(1.62 \pm 1.37) \times 10^{-5}$	-2.27 ± 0.222	0	0	1.9×10^{-5}
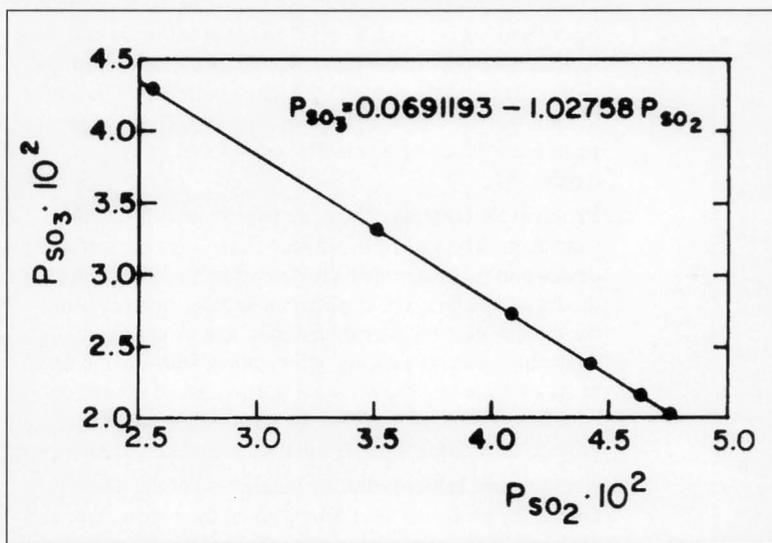
Figure 3. *Linear dependency between p_{SO_3} and p_{SO_2} in the experiments.*

students have increasingly accepted the challenge of being responsible for their work and they now feel more productive and stimulated. As a result, the addition of an economic flavor in the Unit Operations Laboratory has been well received by students who have appreciated its relevance to education leading to a commercial career.

CONCLUSIONS

By providing a balance between an innovative commercial aspect and the traditional engineering-science aspect of the Unit Operations Laboratory, we have introduced elements of "real-life" into laboratory work. Solving problems which have the flavor of industry makes laboratory work more challenging and interesting and, we think, more relevant for the students.

In the future we intend to extend the approach beyond the two experiments discussed here and involve the majority of the Unit Operations Laboratories. Most experiments can be augmented readily, according to the same principles used with our leaching and filtration experiments, by 1) defining profitability equations for an existing plant or process and seeking an economic optimum, or 2) defining a scale-up problem and using laboratory data to predict the economics of a full-scale plant and process.

REFERENCES

1. Coulson, J.M., J.F. Richardson, J.R. Backhurst, and J.H. Harker, *Chemical Engineering. Volume 2: Unit Operations,* 3rd ed., Chapter 9 (Filtration) and Chapter 10 (Leaching); Pergamon Press, Oxford, England (1978)
2. Sinnott, R.K., *Chemical Engineering. Volume 6: An Introduction to Chemical Engineering Design,* Chapter 6 (Costing and Project Evaluation), Pergamon Press, Oxford, England (1983)
3. Davies, W.A., "A Three-Stage Counter Current Leaching Rig for the Senior Laboratory," *Chem. Eng. Ed.,* **23**(2), 96 (1989) ❐

Heterogeneous Reaction Rate Data

Continued from page 25.

Figure 3 shows that p_{SO_3} is a linear function of p_{SO_2}, implying that these variables were actually not changed independently during the experiments. Thus, there is no way to separate the information in the reaction rate data related to these two variables. Plotting p_{O_2} versus p_{SO_3} gives similar results, indicating that there is also a linear dependency between these two variables.

CONCLUSIONS

Initial results have shown that the reaction rate data of Table 1 can be represented well by the rate expression of Eq. (1). It is very tempting to jump to the conclusion that the experimental data verifies the mechanism postulated in Eq. (1), but using numerical and statistical analysis of the data,

we have proven that such a conclusion is completely groundless because

1. The data itself is not experimental, but extrapolated, the accuracy of which is impossible to assess.

2. Because of the large value of the equilibrium coefficient and limited accuracy of the reaction rate values, no effect of the reversibility can be detected in the reaction rate values. As a result, assuming irreversible reaction yields a more accurate correlation over that obtained with the reversible model.

3. The partial pressures of SO_3, SO_2, and O_2 were not varied independently during the experiments; there is linear dependency between the partial pressures of SO_3, SO_2, and O_2. As a result, it is impossible to discriminate between the effects of p_{SO_3}, p_{SO_2}, and p_{O_2} on the reaction rate.

4. The equation

$$r = 1.62 \times 10^{-5} \, p_{SO_3}^{(-2.27)}$$

has been found to best represent the data in Table 1, but because of the limitations of the data that were mentioned earlier and the empirical nature of the power-law rate expression, there is absolutely no certainty that this rate expression is applicable for other combinations of partial pressures.

We have used one particular example to demonstrate several potential pitfalls in correlation of experimental data. It can be expected that in most practical cases not all these pitfalls will show up; but these four points can serve as guidelines in assessing the quality of the data, the accuracy of the correlation, and the adequacy of a model to represent the data.

In conclusion, it is appropriate to quote Churchill, who noted that "...if the observed behavior of the process requires the use of a more complex model than the data justifies, resolution lies in the laboratory rather than in further analysis."

REFERENCES

1. Churchill, S.W., *The Interpretation and Use of Rate Data: The Rate Concept,* revised printing, Hemisphere Publishing Co., Washington, DC (1979)
2. Smith, J.M., *Chemical Engineering Kinetics,* 3rd ed., McGraw-Hill, New York, NY (1981)
3. Olson, R.W., R.W. Schuler, and J.M. Smith, "Catalytic Oxidation of Sulfur Dioxide: Effect of Diffusion," *Chem. Eng. Prog.,* **46**(12), 614 (1950)
4. Shacham, M., and M.B. Cutlip, *POLYMATH 3.0 Users' Manual,* CACHE Corporation, Austin, TX (1993)
5. Brauner, N., and M. Shacham, "Model Discrimination and Error Analysis in Heterogeneous Rate Data Regression for Irreversible Reactions," submitted for publication (1993) ❐