

# Application of Response Surfaces in Evaluating Tool Performance in Metalcutting

Michael R. Delozier

State College, Pennsylvania, USA

Email: n-ball29@comcast.net

Received May 15, 2013; revised June 15, 2013; accepted June 22, 2013

Copyright © 2013 Michael R. Delozier. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## ABSTRACT

This paper advances the collection of statistical methods known as response surface methods as an effective experimental approach for describing and comparing the tool life performance capabilities of metalcutting tools. Example applications presented demonstrate the versatility of the power family of transformations considered by Box and Cox (1964) in modeling tool life behavior as revealed using simple response surface designs. A comparative analysis illustrates a method to gauge the statistical significance of differences in tool life estimates computed from response surface models. Routine use of these methods in experimental tool testing is supported by their ability to produce reliable relative performance representations of competing tools in field applications.

**Keywords:** Response Surfaces; Metalcutting; Tool Life; Power Transformations

## 1. Introduction

Experimental, laboratory-based, tool performance testing in machining operations is an integral part of the product development cycle of metalcutting tools. Ideally, such testing provides information-rich data as a basis for screening various product designs and evaluating potential improvements in existing cutting tool materials, or grades. In an iterative product development process, testing often ultimately focuses on comparing the performance qualities of candidate designs to those of competing grades over a specified application range of cutting, or operating, conditions. Laboratory testing is thus an essential performance review to assure that product design and manufacturing process concepts are correct prior to field trials in actual applications.

The need to compare performance qualities over a range of operating conditions suggests the use of experimental design to efficiently produce the required test data. The purpose of data analysis is to provide insight to potentially important performance differences influencing decisions to pursue further development or field-testing.

A performance quality measure frequently analyzed in metalcutting applications is tool life. As varied as the field applications themselves, so are the criteria used to determine the end of life of cutting tools in these applications. In the laboratory, a definition of tool life providing

a common basis for comparing tool performance capabilities is required for useful analysis to proceed. The definition ordinarily used is the cutting time until the wear on the tool reaches a pre-specified level, or the tool catastrophically fails. Depending on test objectives, other quality measures such as the surface finish of the material being machined (the “workpiece” material), the ability to control chip formation during machining, power consumed, and the magnitude of various cutting forces may be analyzed as well.

This paper presents a statistical approach to describe and compare the tool life capabilities of metalcutting tools using response surface methods, a coordinated system of experimental design, regression analysis, and graphic presentation. A mathematical approximation model, often termed a response surface model, expressing tool life in terms of selected machining variables, whose settings define the range of cutting conditions of interest, is developed using test data from a designed experiment. Examining the model using contour plots provides a simple analysis of tool performance capabilities with respect to changes in these variables.

## 2. Empirical Models of Tool Life Behavior

Consideration of the dominant features of tool life data is basic to proper model formulation. Over sufficiently wide ranges of cutting conditions, tool life tends to ex-

hibit non-linearity of relationship and changing variability as illustrated in **Figure 1** adapted from Fung [1]. Power transformations of the response are often effective in modeling such behavior. Therefore, we tentatively assume that the relationship between tool life,  $y$ , and  $k$  independent machining variables,  $\xi_1, \xi_2, \dots, \xi_k$ , can be adequately represented by

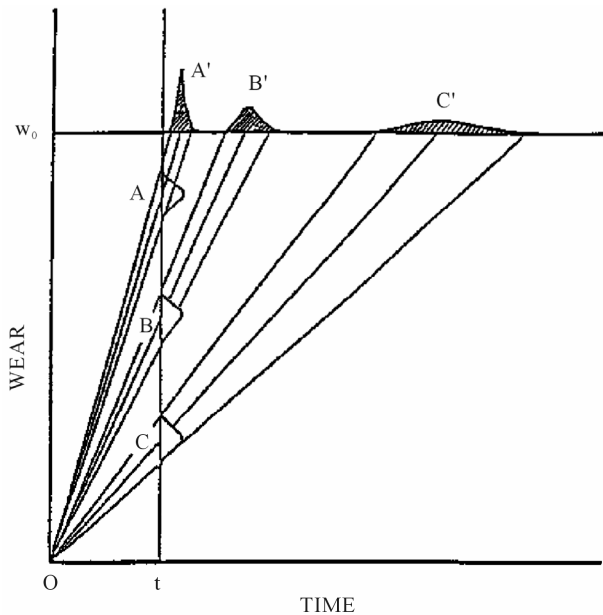
$$y_i^{(\lambda)} = g(\xi_i, \theta) + \varepsilon_i, \quad (1)$$

where  $y^{(\lambda)} = y^\lambda$  if  $\lambda \neq 0$ ,  $y^{(\lambda)} = \ln(y)$  if  $\lambda = 0$ ,  $\xi_i$  is a vector of independent variable settings determining the  $i^{\text{th}}$  experimental test condition,  $\theta$  is a vector of parameters to be estimated from data,  $g$  is a low-order polynomial in the  $\xi$ 's, and the  $\varepsilon_i$  are statistical errors that follow, at least approximately, the usual linear model assumptions.

As noted in Balakrishnan and DeVries [2], many researchers have advocated the use of a linearized tool life model of the form

$$\ln(y_i) = \theta_0 + \sum_j \theta_j \ln(\xi_{ji}) + \varepsilon_i, \quad (2)$$

or an extension including second-order terms in the logarithms of the  $\xi$ 's. The model in (2) is a linearized generalization of the tool life-cutting speed relationship proposed by F. W. Taylor for high-speed steel cutting tool materials in the early 1900s. While in some applications these models may be useful, there does not seem to be any particular reason why model adequacy and simplicity of relationship will always be achieved in the logarithms. The form given in (1) thus provides a flexible alternative starting point for the modeling process.



**Figure 1.** Schematic of tool life variability at three cutting conditions with an end of life wear criterion  $w_0$ .

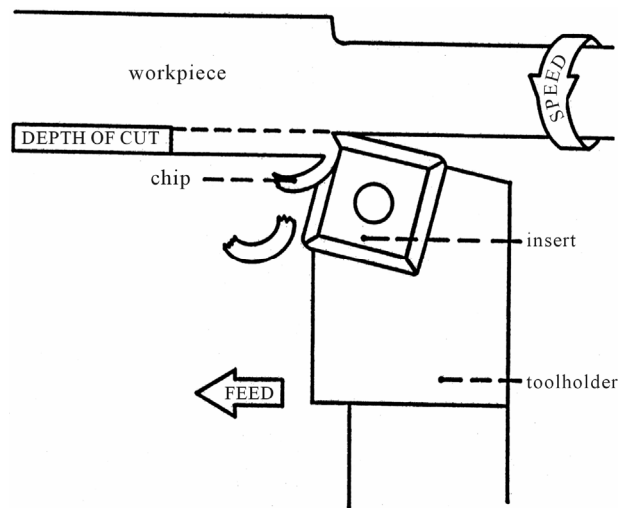
### 3. Applications

#### 3.1. Application 1: A Power Family Model of Tool Life

##### 3.1.1. Description of Application and Experimental Design

Metalcutting tools are designed and manufactured in various sizes and shapes to service a wide range of machining applications. In addition, various geometric “chip control” designs may be shaped on the tool surface in manufacturing to control workpiece chip formation during cutting operations, and to enhance tool life. This application is part of a larger evaluation aimed at characterizing the effects of three variables, cutting speed, feed rate, and depth of cut, on the tool life performance of tools produced with certain chip control designs.

The tool life data given in **Table 1** are the results of a completely randomized experiment for a cutting tool grade manufactured with a commonly used chip control design. The machining operation was turning a medium-carbon steel workpiece material on a lathe (**Figure 2**). A central composite test design with the axial points located at the center of each face of the unit cube (**Figure 3**) was used to produce the required tool life data. A desirable characteristic of this design is its ability to estimate higher-order effects with only three levels for each factor. Also, this design provides protection from having features of fitted models strongly influenced by one or a few test results remotely positioned in the factor space—as may occur say in using a rotatable central composite design having unreplicated axial points located outside the unit cube. Replications at locations other than the design center were run to reveal the approximate variance profile across the operating range, and so obtain more realistic estimates of tool life uncertainty. As is common in practice, the levels of the three factors are coded giving

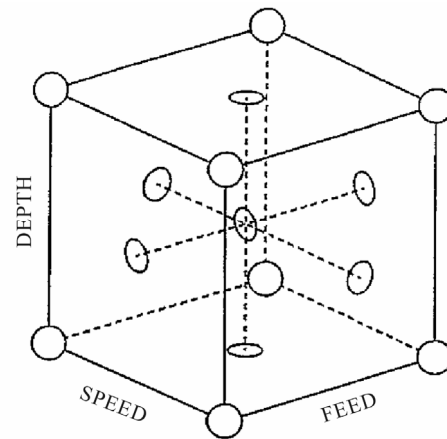


**Figure 2.** Schematic of a turning operation in machining.

**Table 1. Cutting conditions and test results.**

Speed (sfm)	Feed (ipr)	Depth of Cut (in)	Tool Life (min)
650	0.010	0.050	49.4
650	0.010	0.050	31.5
650	0.010	0.050	50.0
800	0.010	0.050	22.0
800	0.010	0.050	30.0
800	0.010	0.050	31.9
650	0.026	0.050	10.0
650	0.026	0.050	8.1
650	0.026	0.050	6.5
800	0.026	0.050	3.2
800	0.026	0.050	1.4
800	0.026	0.050	2.3
650	0.010	0.200	32.2
650	0.010	0.200	29.1
650	0.010	0.200	40.2
800	0.010	0.200	13.5
800	0.010	0.200	14.1
800	0.010	0.200	14.6
650	0.026	0.200	1.4
650	0.026	0.200	1.5
650	0.026	0.200	1.5
800	0.026	0.200	0.5
800	0.026	0.200	1.1
800	0.026	0.200	0.6
650	0.018	0.125	9.1
800	0.018	0.125	3.0
725	0.010	0.125	27.5
725	0.026	0.125	2.9
725	0.018	0.050	12.3
725	0.018	0.200	4.2
725	0.018	0.125	7.1
725	0.018	0.125	8.6
725	0.018	0.125	4.2

Note: Speed is in units of surface feet per minute (sfm), feed in inches per revolution (ipr), depth of cut in inches, and tool life in minutes. Tool life is determined by the first occurrence of 0.015 inch uniform flank wear, 0.004 inch crater depth, 0.030 inch localized wear, or catastrophic failure.



**Figure 3. Central composite design used in Application 1.**

as predictors  $x_1 = (\text{speed} - 725)/75$ ,  $x_2 = (\text{feed} - 0.018)/0.008$ , and  $x_3 = (\text{depth} - 0.125)/0.075$ .

**3.1.2. Response Surface Model Selection**

The adequacy of the model

$$\widehat{y}^{0.2} = 1.4479 - 0.1270x_1 - 0.3823x_2 - 0.1308x_3 - 0.0416x_2x_3 + 0.1143x_2^2 + 0.0431x_1x_2x_3 \quad (3)$$

is easily established. The “hat” notation is used to denote predicted values determined by the estimated regression equation.

A normal probability plot of the residuals and a plot of the residuals against the fitted values show the success of transformation as a remedy for error pattern inadequacies characteristic of fits in the original response scale. The estimated coefficients are highly significant (maximum p-value for the individual t-tests of significance is 0.007) and a large proportion of the total variation in the transformed tool life values is explained by the fit ( $R^2$  is 0.975). Lack of fit is not indicated by either the pure error test or the Minitab Statistical Software data subsetting test [3].

In this application, an initial model screening was carried out using stepwise variable selection for chosen  $\lambda$  over the interval  $-1$  to  $+1$ . Subsequent ranking of various models terminating the stepwise algorithm was based on computing

$$\text{PRESS} = \sum_i (y_i - \hat{y}_{-i})^2 \quad (4)$$

Here,  $\hat{y}_{-i}$  is the predicted value of  $y_i$  from a fit using all of the data except the  $i^{\text{th}}$  case (the  $i^{\text{th}}$  tool life observation and its associated independent variable settings). Better models have relatively small values of PRESS. Allen [4] gives a computational form for PRESS easily adapted to the class of models considered in (1).

From this heuristic evaluation, several candidate models were selected for further analysis to evaluate ade-

quacy of fit. There appears to be little difference between response scales regarding simplicity of relationship for these data. The need to include interactions and, in some cases, a quadratic effect in models to describe tool life behavior was evident for all transformations tried. The model in (3) has a PRESS of 649, which is slightly larger than the smallest found (a model with the square root of tool life as the response with a PRESS of 641 was found). The model given in (3) was ultimately chosen over that with the slightly smaller PRESS as a result of its somewhat more pleasing residual patterns.

Contour plots (not included) displayed in the original units of the response show the loci of cutting conditions giving specified estimates of the median of the predictive distribution of tool life. Overlaying such contour plots generated from the approximation models of several chip control designs provides a simple means to simultaneously compare performance characteristics.

### 3.2. Application 2: A Comparative Tool Life Analysis

#### 3.2.1. Description of Application and Experimental Design

This application is part of a larger metalcutting product performance evaluation. The test objective was to obtain a tool life comparison of two competing cutting tool grades, Grade A and Grade B, in turning a medium-carbon steel workpiece material on a lathe. Tests for each grade were set up using a  $2^2$  factorial plus center point design with the factors cutting speed and feed rate. The test runs were collectively randomized and the tool life for each grade recorded as shown in **Table 2**. In this application, the levels of the factors are coded giving as predictors  $x_1 = (\text{speed} - 650)/150$  and  $x_2 = (\text{feed} - 0.021)/0.006$ .

#### 3.2.2. Preliminary Remarks

Inspection of the test data suggests that statistically (and likely practically as well) meaningful differences in tool life level favoring Grade A may exist at and about the center of the factor space. Without formal statistical treatment of the data, a benefit of using the experimental design is immediately realized. That is, important relative performance information may have remained concealed if testing was limited exclusively to say near the center, or near an extreme, of the factor space.

#### 3.2.3. Response Surface Model Selection

For each respective tool grade, ranking of the first-order and first-order plus interaction fits based on PRESS for chosen  $\lambda$  over the interval  $-1$  to  $+1$  was carried out to provide guidance in model selection. A complete listing of computational results is not given. To summarize, the models found to have the smallest PRESS for Grade A

**Table 2. Cutting conditions and test results.**

Speed (sfm)	Feed (ipr)	Tool Life (min)	
		Grade A	Grade B
500	0.015	67.0	84.4
500	0.015	101.9	91.2
500	0.015	63.6	66.7
800	0.015	23.5	16.0
800	0.015	17.6	15.2
800	0.015	21.3	17.6
500	0.027	17.9	24.6
500	0.027	25.3	15.3
500	0.027	25.4	30.4
800	0.027	0.4	1.1
800	0.027	0.6	0.5
800	0.027	0.5	0.9
650	0.021	21.4	11.8
650	0.021	19.2	8.9
650	0.021	22.6	10.6

Note: Speed is in units of surface feet per minute (sfm), feed in inches per revolution (ipr), and tool life in minutes. Tool life is determined by the first occurrence of 0.015 inch uniform flank wear, 0.004 inch crater depth, 0.030 inch localized wear, or catastrophic failure.

and Grade B respectively are the first-order fit with  $\lambda$  of 0.5 (PRESS of 1607) and the first-order plus interaction fit with  $\lambda$  of  $-0.1$  (PRESS of 983). However, the PRESS value for the first-order plus interaction fit for Grade B with  $\lambda$  of 0 is not very different from the smallest found. Thus, the suitability of this fit as a model of the tool life performance for Grade B is examined.

A likelihood-based procedure for estimating  $\lambda$  from the data is given by Box and Cox [5]. Its application via a Minitab Statistical Software macro for the first-order and first-order plus interaction model forms for Grades A and B respectively indicates general agreement with the rankings based on PRESS. Moreover, in the case of Grade B, tests for lack of fit suggest that the interaction term is not removable by transformation.

A normal probability plot of the residuals and a plot of the residuals against the fitted values for the first-order plus interaction fit with the logarithm of Grade B tool life as the response indicate no model inadequacy. These diagnostic plots for the first-order fit with the square root of Grade A tool life as the response are not as satisfying, seemingly due in part to the effect of the largest observation. However, to avoid suppressing variation at a condition expected to show sizable response variation, we re-

tain this case and select a model otherwise fitting the data well. The first-order model with the square root of tool life as the response appears to fit well and is selected. Soothing the decision to retain this observation is that a later test at this condition resulted in a tool life of nearly two hours.

A comparative performance analysis will thus be based on the following models for Grade A and Grade B respectively:

$$\widehat{y}_A^{0.5} = 4.6726 - 2.0653x_1 - 1.9582x_2 \quad (5)$$

$$\widehat{\ln(y_B)} = 2.4778 - 1.2363x_1 - 1.0724x_2 - 0.4384x_1x_2. \quad (6)$$

### 3.2.4. A Comparative Analysis Method

A simple comparative analysis may be carried out by superimposing contours of the estimated response surfaces and observing the relative position of contours of equal estimated median tool life. **Figure 4** shows the superimposed contours of 10, 30, and 50 minutes median life. The apparent effect that Grade A is capable of operating at higher and more productive cutting conditions, while yielding the same median life as Grade B, is most prominent near the center of the factor space. That is, meaningful performance differences favoring Grade A are likely to exist in this region. However, without an indication of the variability associated with these estimates, differences that are statistically important, or significant, are not discernible.

An ad hoc, but useful analysis approach that incorporates the variability of model estimates in comparing performance differences is to form the surface that is the difference of the two tool life models standardized by a measure of uncertainty. A contour plot of this surface is useful in identifying regions of test conditions where the differences in estimated median response are large relative to uncertainty.

A slight complication occurs in cases such as this where the models are fit using different response transformations. Sensible choices of a common scale in which to compare differences include that of either model, or the original response scale. In any case, a first-order propagation of error approximation can be used to estimate variability in an alternate scale.

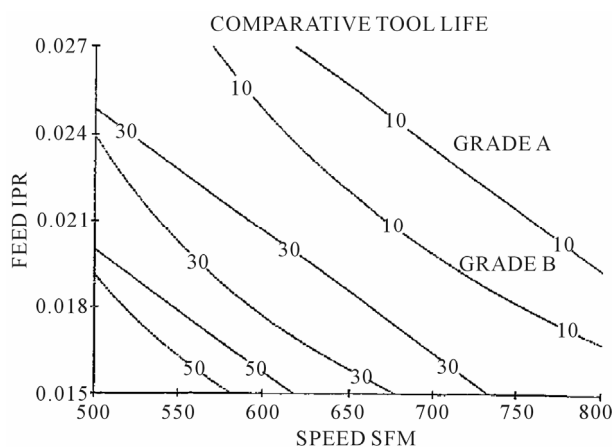
In this application, an analysis in the logarithmic scale may be carried out by plotting contours of the surface (**Figure 5**)

$$\delta(x_1, x_2) = \left\{ \left[ 2\ln(\widehat{y}_A^{0.5}) \right] - \widehat{\ln(y_B)} \right\} / (V_A + V_B)^{0.5}, \quad (7)$$

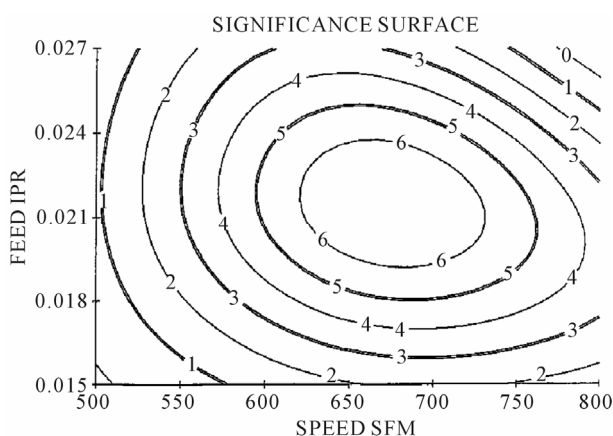
where  $V_A = \left[ 4 / (\widehat{y}_A^{0.5})^2 \right] \widehat{\text{Var}}(\widehat{y}_A^{0.5})$  and

$V_B = \widehat{\text{Var}}[\widehat{\ln(y_B)}]$ .  $\widehat{\text{Var}}$  denotes the estimated variance of the indicated argument.

Loosely speaking, we shall say statistically significant



**Figure 4.** Selected contours of the estimated tool life surfaces for the cutting tool grades in Application 2.



**Figure 5.** Contours of the estimated significance surface.

performance differences exist at test conditions where this surface is sufficiently high or low. In practice, conditions with  $|\delta| \geq 3$  have been found to adequately approximate those exhibiting meaningful tool life differences in similar field applications. The egg-shaped region about the center of the factor space shown in **Figure 6** has  $\delta \geq 3$  indicating a tool life advantage for Grade A. It is reassuring to find that analysis in either the square root or the original scale identifies essentially the same region of tool life advantage for Grade A (**Figure 7**).

Extension of this analysis to more than two independent variables is straightforward though visualizing the results would require multiple contour plots, each generated with the variables chosen to be “off-axis” fixed at desired settings.

### 3.3. Application 3: Adaptability to a More Complex Testing Situation

In some situations where appreciable systematic variation in the test environment is expected, test methods may be suitably modified to yield useful performance

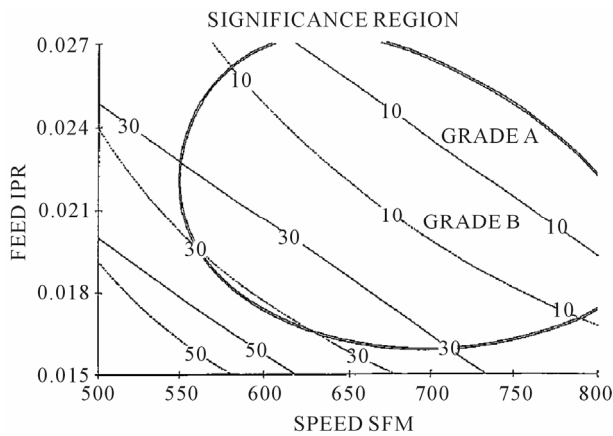


Figure 6. Significance region superimposed on the contours of the estimated tool life surfaces.

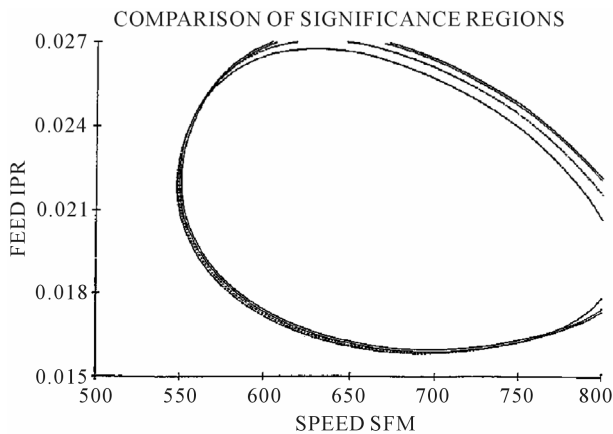


Figure 7. Significance regions for comparisons in the original, square root, and logarithmic scales.

information. For example, consider a situation where a tool life comparison of two grades is desired. Expecting substantial variation in workpiece material properties (e.g., hardness) over the course of testing, consider generating tool life data as follows: a cutting edge of one of the two grades is run for a predetermined “short” cutting time, using a machining condition randomly selected from those prescribed in the experimental design, and tool wear measurements made; an edge of the competing grade is run for the same time and wear measurements made; the edge first run is run again for the same time and wear measurements made; the competing edge is run again for the same time and wear measurements made; and so on until the end of life is reached for both grades. This procedure is repeated at each respective condition. Testing in this “back-to-back” manner creates a paired data structure. Actual tool life data generated in this fashion are given in **Table 3**. In this data set, the cutting edges resulting in 89.1 and 29.6 minutes tool life were run together, 56.9 and 20.5 were run together, 8.4 and 3.5 were run together, and so on.

Table 3. Cutting conditions and test results.

Speed (sfm)	Feed (ipr)	Tool Life (min)	
		Grade C	Grade D
400	0.015	89.1	29.6
400	0.015	56.9	20.5
700	0.015	8.4	3.5
700	0.015	10.2	4.3
700	0.015	5.1	2.5
400	0.025	34.0	19.0
400	0.025	24.0	20.1
400	0.025	22.0	18.2
400	0.025	24.5	15.6
700	0.025	1.9	1.3
700	0.025	1.6	2.0
700	0.025	1.5	1.1
550	0.020	10.8	7.6
550	0.020	11.2	2.7
550	0.020	8.7	5.6
550	0.020	12.0	6.6

Note: Speed is in units of surface feet per minute (sfm), feed in inches per revolution (ipr), and tool life in minutes. Tool life is determined by the first occurrence of 0.015 inch uniform flank wear, 0.004 inch crater depth, 0.030 inch localized wear, or catastrophic failure.

For brevity, an analysis of the data in **Table 3** is not pursued. An analysis approach that has generally provided satisfactory results in practice involves developing an approximation model using the differences between paired observations as the response values. Estimated differences computed from this model, standardized by a measure of uncertainty, can then be used to assess statistical significance over the operating range covered by the test.

#### 4. Summary

This paper advances response surface methodology as an effective experimental approach for describing and comparing the tool life performance capabilities of metalcutting tools. Such an approach provides a means of identifying important performance differences between the tools tested over a specified range of operating conditions. Several example applications demonstrate the versatility of the power family of transformations in modeling tool life behavior as revealed using simple response surface designs. A comparative analysis application illustrates a method to gauge the statistical significance of differences in tool life estimates computed from response

surface models. Therefore, test conditions producing statistically important differences may be identified thus approximating operating regions of field performance strength or weakness. Routine use of these methods in experimental tool testing is supported by their ability to produce reliable relative performance representations of competing tools in field applications.

## 5. Acknowledgements

The tool performance modeling and analysis methodology presented in this paper was contributed by this author, now retired, while a statistician supporting product research and development, and process improvement, for a leading manufacturer and supplier of metalworking tools and tooling systems. The author gratefully acknowledges this former employer for the opportunity to develop and contribute this work, and as the source of the schematic displayed in **Figure 2**.

## REFERENCES

- [1] C. A. Fung, "Statistical Topics in Off-Line Quality Control," Ph.D. Thesis, University of Wisconsin—Madison, 1986, p. 167.
- [2] P. Balakrishnan and M. F. DeVries, "Analysis of Mathematical Model Building Techniques Adaptable to Machinability Database Systems," *Eleventh North American Manufacturing Research Conference Proceedings, Society of Manufacturing Engineers*, 1983, pp. 466-475.
- [3] Minitab, Inc., "Minitab User's Guide 2: Data Analysis and Quality Tools—Release 12," 1998.
- [4] D. M. Allen, "The Prediction Sum of Squares as a Criterion for Selecting Predictor Variables," Technical Report Number 23, Department of Statistics, University of Kentucky, 1971.
- [5] G. E. P. Box and D. R. Cox, "An Analysis of Transformations (with discussion)," *Journal of the Royal Statistical Society, Series B*, Vol. 26, 1964, pp. 211-252.