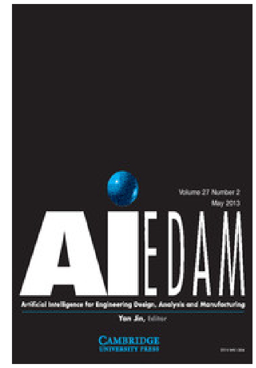


AI EDAM
Artificial Intelligence for Engineering Design, Analysis and Manufacturing
CAMBRIDGE UNIVERSITY PRESS

Artificial Intelligence for Engineering Design, Analysis and Manufacturing (2000), *14*, 181–192. Printed in the USA.

Use of graphic displays of process capability data to facilitate producibility analyses

ANNA C. THORNTON AND MELISSA TATA
Mechanical Engineering Department, Massachusetts Institute of Technology, Cambridge, MA

(RECEIVED August 6, 1999; ACCEPTED March 24, 2000)

Abstract

To specify realistic tolerances and ensure acceptable product yields, a product development team must take into account the current process capability of the factory. Using historical process capability data in product development is frequently discussed in academic and industry literature. However, while many organizations have databases, the use of historical data in design is limited. This paper addresses the technical challenges of implementing a producibility analysis tool. First, a description of how data is currently collected and recorded is given. Second, the infeasibility of a "just-press-a-button" system for checking producibility is described. Finally, a new paradigm for data access is proposed, implemented in software, and tested on data used by industry.

Keywords: Process Capability Data; Producibility Analyses; Iterative Query/Visualization Method

1. INTRODUCTION

Producible—To be able to give being, form, or shape to.

Dramatic improvement in product cost and quality can be achieved by ensuring a design is producible before transitioning it to production. Ensuring producibility reduces the need for rework and/or late design changes. There are many aspects of producibility, but this paper focuses specifically on the ability of a process to meet target tolerances. To verify tolerances prior to production, many companies are building process capability databases. These databases contain measurements taken during production. The ideal system integrates the process capability database with computer aided design (CAD) and variation modeling systems. The producibility is checked using a "just-press-the-button" method in which the database is automatically queried, the tolerances evaluated, and/or variation models populated. While the use of historical process capability data appears to be a relatively simple proposition, in reality, it is a complex problem.

To understand how companies are currently using process capability data, the authors surveyed more than twenty large design and manufacturing firms (Tata & Thornton, 1999). All but one of the design and manufacturing companies we interviewed have invested significant resources to develop, implement, and populate process capability databases. However, only a small fraction of product development teams in these organizations use the data to evaluate producibility.

The survey identified organizational and technical barriers to process capability data use. Organizational barriers include not having systematic procedures and policies for using the data (e.g., incentive structures and standard design practices) and not having good communication between design and manufacturing (e.g., teams did not trust each other, limited access to databases, and poor communication of needs). Technical barriers include poor indexing schemes, lack of statistical validity, excessive time required to access the data, and poor user interfaces. In general, the technical barriers prevent designers from getting the right data quickly.

First, a detailed description of how data is currently collected and recorded is given. Second, the reasons why a "just-press-a-button" system is infeasible are given. Finally, a new paradigm for data access is proposed and illustrated using a data set from industry.

2. LITERATURE

Understanding the impact of variation on product performance and cost is critical in product development. Several

Reprint requests to: Anna C. Thornton, Mechanical Engineering Department, 3-449 MIT, 77 Mass Ave., Cambridge, MA 02139. E-mail: acthornt@mit.edu

articles (Parkinson et al., 1993; Chase et al., 1996; DeGarmo et al., 1997) stress the importance of tolerance allocation. If tolerances are too tight, unnecessarily expensive operations are used and if tolerances are too loose, the part may not function properly. Tolerances should be optimized to reduce mechanical errors (Lee et al., 1993; Lin et al., 1997; Zhang & Wang, 1998), minimize assembly problems (Ting & Long, 1996), and improve product performance (Wang & Ozsoy, 1993; Michelena & Agogino, 1994). Setting tolerances to match process capability and design intent is also the subject of significant literature (Liu et al., 1996; Srinivasan et al., 1996; Gao et al., 1998).

Several papers have proposed methods to predict process variation using process models (Kazmer et al., 1996; Zemel & Otto, 1996; Frey et al., 1998; Soyucayli & Otto, 1998; Yu & Ishii, 1994). Creating process models is time consuming and may not be feasible for all processes. Several articles discuss using process capability data in the product delivery process. For example, Naish (1996) describes the role process engineers play in selecting processes capable of meeting target tolerances. Several articles address the specific problem of using process capability in electronic systems design (Lucca et al., 1995; Nagler, 1996). Campbell and Bernie (1996) discuss requirements for a formalized rapid prototyping database. Perzyk and Meftah (1998) describe a process selection system that includes general data on process capabilities. Baldwin and Chung (1995) discuss some methods for managing vast quantities of data using a classification hierarchy.

There is a large amount of literature on the subject of information visualization and retrieval. There are several journals and conferences dedicated to the subject. The relevant literature will be reviewed throughout the paper.

In summary, while many papers discuss the possibility of using process capability in design using an automated checking system, none describe the implementation of a working system. The authors believe the ideal state is not possible with current data and technology—*the problem is just not that simple*. Rather, methods that effectively use currently available data need to be developed.

3. CURRENT STATE OF PROCESS CAPABILITY DATABASES

All of the companies the authors interviewed used similar database structures and indexing methods (Tata & Thornton, 1999). This section describes the typical database structure, indexing and retrieval methods. The summary analyzes the limitations imposed by the database structure.

3.1. Process capability data

Process capability is the amount of variation introduced when a feature is created by a process. In most cases, the variation is assumed to have a Gaussian distribution, and the standard deviation and bias are used to characterize the process. In addition, product development teams often use *Cpk* and *Cp*, the normalized ability to hold a tolerance, to evaluate producibility. Typically, the team aims to have all tolerances produced with a *Cpk* of 1.33.

The standard deviation and bias for a process are estimated from the individual data points. The run's bias, b, is a function of n individual data points, x_i and the target value, m.

$$\mu = \sum_{i=1}^{n} \frac{x_i}{n} \quad \text{and} \quad b = \mu - m \tag{1}$$

The run's standard deviation is a function of the individual points, x_i, and the mean μ.

$$\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{n-1}} \tag{2}$$

Cp and *Cpk* are a function of the upper and lower tolerance limits $[LL, UL]$.

$$Cp = \frac{UL - LL}{6\sigma} \quad \text{and} \quad Cpk = \frac{\min(UL - \mu, \mu - LL)}{3\sigma} \tag{3}$$

3.2. Influences on process capability

When a team asks the question, *what is the standard deviation and mean shift for hole-drilling*, they often get back the answer *it depends*. The search for the correct answer is non-trivial because many factors other than the feature type can influence the final process variation.

- *Process*: Holes can be created using a variety of processes. Each process will introduce different amounts of variation.
- *Material*: Given the same process, some materials will introduce more variation than others will.
- *Feature geometry*: The process variation for hole-drilling will be different depending on what aspect of the feature you are measuring. For example, the hole depth may have less variation than the diameter.
- *Feature size*: The process variation in large diameter holes may be larger than that for small diameter holes.
- *Machine*: In many cases, multiple machines are capable of producing the same part. Many designers will tolerance assuming the best process capability. However, the best machine may not have the capacity to produce all parts. Consequently, overflow parts may be made on older, less capable machines.
- *Tolerance*: The tolerance on the drawing may change the process used and hence the variation introduced. For example, for tight tolerance surfaces, NC path planners may cut at slower speeds. However, tighter toler-

Table 1. *Sample record*

Feature Code	Target	UL	LL	Machine	Measurement
M1.1 P7.3.3 F1.1.1	0.0766	0.0966	0.0666	1031	0.0756

ances typically come at a cost: larger labor content, processing cost, and/or throughput time.

- *Operator*: The operators can influence process variability. Some operators are better at coaxing the machine than others.
- *Part characteristics*: The location of a feature in a part can influence process variation. For example, more variation will be introduced when drilling a hole in a thin part than in a thick part. Fixtures and postprocessing will also influence the final variation.
- *Number of parts*: In low-volume production, parts may be made by specialists in a job-shop environment. This may or may not increase quality. Using automation in high-volume may standardize processes but the job-shop may be more capable of controlling variation.

In most cases, these factors cannot be treated as separate effects; for example, the material and process jointly determine the process variation. Many process characteristics are known by the designer and some are not. For example, the designer will know the feature type, dimensions, materials, and tolerance. However, they may not know what machine or process will be used to produce the part or which operator will make the part. Some characteristics are easily encoded (i.e., the volume or material) while others may be more difficult or impossible to encode (i.e., geometric characteristics).

3.3. Database structures

A process capability database is typically used to store the measurements taken in the factory during production. Typically, each measurement is entered in the database as a single *record*. The process characteristics are recorded along with the measurement data. Typically, only the easily encoded influences are included: that is, part number, feature code, date, operator, machine, special causes of variation, target nominal, and target tolerance.

Table 1 shows a portion of a sample record. All of the data shown in this paper is extracted from an actual industry database. The numbers and codes have been adjusted to mask any proprietary data. In this case, the feature code, machining area, target value, upper and lower limit, and machine number are used to characterize the measurement. The measurement value was 0.001 below the target value (within the allowable range).

Data is then grouped into *runs*. The records in a run share the same process characteristics—that is, part number, feature code, operator, machine, nominal and tolerance. Table 2 shows the run containing the record in Table 1. The characteristics for each run are calculated and can include n, the number of records, b, bias, σ, the standard deviation, Cp, and Cpk.

3.4. Feature codes

One of two methods is typically used to index a database: drawing-based or feature-based. Drawing-based systems are indexed by the part/drawing and feature number. Older databases are typically indexed in this manner. This method is ideal for monitoring part quality, but it cannot be easily used during design because the team may not be familiar with older parts and their part numbers. In addition, older drawing-based systems may not have appropriate critical features selected or monitored.

The design team needs to find the runs that most closely match their design using the information they have and the information provided by the database. They may want to tolerance a hole diameter or the relative location between two holes. In addition, they often know the material and the machining process being used. To enable a user to quickly look up data, many databases are being indexed using a feature code.

A feature code is a systematic method for describing typical part/process/material combinations. The standard features, materials, and processes are given unique codes. A

Table 2. *Sample run*

Feature Code	Target	*LL*	*UL*	Machine	Bias	Std dev	*n*	*Cp*	*Cpk*
M1.1 P7.3.3 F1.1.1	0.0766	.0966	0.0666	1031	0.0001	0.0047	136	1.07	0.72

Table 3. *Sample feature codes*

Material	Process	Feature
4. Titanium	7. Hole Preparation	3. Hole
4.1. Extrusion	7.1. Drill	3.1. Location
4.1.1 6AL6V	7.1.1. Handheld	3.2. Length
4.1.2 6AL6V2SN	7.1.2. Semi-Automatic	3.3. Diameter
4.1.3 6AL4V	7.1.3. Numerical Control	3.4. True Position
4.2. Plate, Sheet	7.2. Countersink	3.5. Perpendicularity
4.2.1 6AL6V		
4.2.2 6AL6V2SN		
4.2.3 6AL4V		

combination of a feature, material, and process describes each data record. Typically, each indexing code has a hierarchical structure used to facilitate coding and retrieval. Table 3 shows a sample coding from an existing indexing system. If a machinist measures the length of a hole made using a handheld drill in a Titanium extrusion 6AL6V the measurement is recorded using the code M4.1.1. P7.1.1 F3.2

Using a feature-indexing scheme enables the team to find data without having to know the drawing numbering system. The common classification scheme also helps the sharing of information between manufacturing groups. However, using indexing is not without problems. One indexing scheme we saw could encode 52 million possible combinations but only 50,000 of the index combinations (0.1%) were feasible. In addition, many feasible index combinations were not populated. The indexing scheme must be developed such that the specificity of the coding is appropriate. Too detailed codes will result in a sparse dataset. Too general codes may be difficult to interpret.

3.5. Access methods

Design teams can access process capability data through either indirect or direct methods. The first requires a formal request to a capability expert who interrogates the database and returns a process capability report. In the second, the user directly accesses the database, but may request assistance from a capability expert to interpret the data.

3.5.1. Indirect data request

In this approach, a specific functional group is responsible for maintaining and accessing the database. Designers typically request a process capability analysis for a specific feature. A process expert then queries the database using their own process knowledge and returns the expected standard deviation and bias. The process capability experts are process engineers and/or quality experts who have detailed knowledge of what data is statistically valid, what parts are in the database, and what alternative data to use when the feature code is unpopulated. The amount of time required to turn around a process verification request can vary from a day to a week. This approach results in accurate assessments, but the process is time consuming. It is typically used only when tight tolerances are required and the cost of not achieving the tolerance is expensive.

3.5.2. Direct access

To answer the question "what is the process capability for a feature," it is necessary to find the run most closely resembling the feature. Typically, databases have a query window similar to Figure 1. This type of query structure has two problems. First, the user may not know some codes; for example, they may not know the operator. This may encourage the user to hunt the database until they find a record that gives them the answer they want. Next, if the database is sparsely populated, the user must adjust the query until a populated run is found.

When the designer finds a run, the computer generates an output similar to Figure 2. The output typically includes a histogram, the input parameters, and the statistics. The lines on the histogram represent the upper and lower tolerance limits.

If a single run cannot be identified that exactly matches the current feature, the user may choose to look at several runs simultaneously. Figure 3 shows the data for the same target feature code and tolerance width for two different dimensions: 0.0766 and 0.0866.

Where there are many runs, simultaneously viewing and interpreting a set of charts can be difficult. In this case, the user may be given a table of data (Table 4).

Finally, the user may be tempted to average the runs together. When all runs in Table 4 are averaged the data in

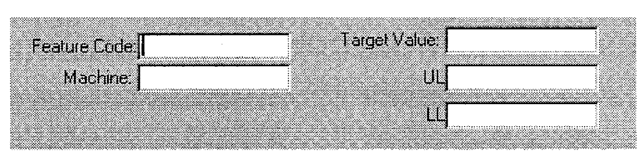
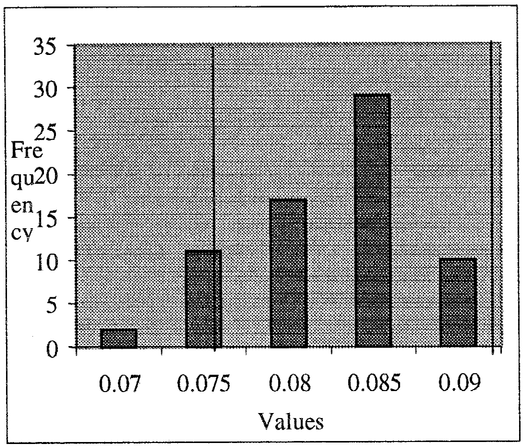


Fig. 1. Sample query form.



Feature Code	Target	UL	LL	Machine
M4.1.1 P7.1.1 F3.2	0.0866	0.1066	0.0766	1031

Number	Std dev	Mean	Cp	Cpk
72	0.0048	0.0856	1.04	0.62

Fig. 2. Single histogram and data.

Table 5 results. In this case, a tolerance bound of [−0.01, +0.25] is assumed. However, it is impossible to guarantee the statistical validity of the result because the standard deviations and biases are likely to be from different populations. When compared to Table 4, it is obvious that the average is not representative of the capability of the individual runs.

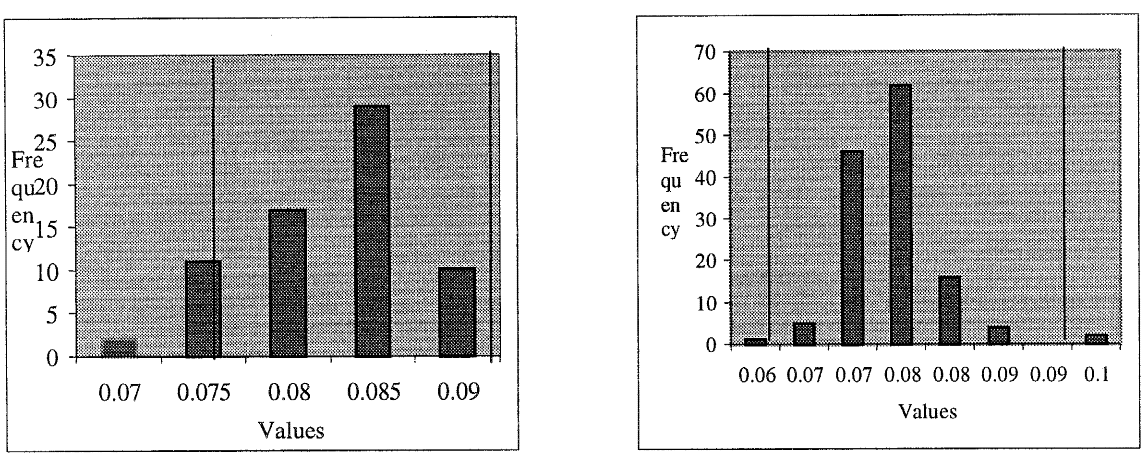
3.6. Summary

At the start of the paper, the hypothesis was put forth that a "just-press-the-button-and-check-the-tolerances" system was not feasible. Because the data contained in the database is historical data, it is not possible to exactly determine if variation will exceed allowable limits on a new part. It is only possible to access a set of possible candidate runs and determine if there is a good chance that the variation will be acceptable. In many cases, some runs will indicate acceptability and others will not. In many cases, the process characteristics that influence the variation are either not known or cannot be encoded.

The following section proposes a scheme for searching process capability databases that uses a combination of statistics, visualization tools, and broad categorization of acceptability to enable the design team to rapidly assess the producibility of a tolerance.

4. ITERATIVE QUERY/VISUALIZATION METHOD

We propose an iterative process of process capability analysis that uses a visualization method to facilitate the search. Shneiderman (1997), in his book on designing user interfaces, describes a four-step process to access and analyzes complex data from a database: formulation, action, results, and refinement. Formulation is the method used to select a subset of the database. Action is the process by which the database is queried. In this case, the action in relatively simple as it is an SQL query. The visualizations of the data are the results. Finally, refinement is the process by which the



Feature Code	Target	UL	LL	Machine
M4.1.1 P7.1.1 F3.2	0.0866	0.1066	0.0766	1031

Number	Std dev	Mean	Cp	Cpk
72	0.0048	0.0856	1.04	0.62

Feature Code	Target	UL	LL	Machine
M4.1.1 P7.1.1 F3.2	0.0766	0.0966	0.0666	1031

Number	Std dev	Mean	Cp	Cpk
136	0.0047	0.0767	1.07	0.72

Fig. 3. Multiple histograms and data.

Table 4. *Output table*

Target	*LL* min	*UL* min	Machine	Bias	Std dev	*n*	*Cp*	*Cpk*
0.2466	−0.01	0.025	1145	0.0046	0.0043	50	1.35	1.13
0.2466	−0.01	0.025	1032	0.0053	0.0059	42	0.99	0.86
0.2466	−0.01	0.025	1031	0.0033	0.0023	6	2.49	1.9
0.2266	−0.01	0.025	1145	0.0059	0.0091	56	0.64	0.59
0.2266	−0.01	0.025	1032	0.0072	0.0104	219	0.56	0.55
0.2266	−0.01	0.025	1031	0.0022	0.0029	6	1.99	1.39
0.2066	−0.01	0.025	1145	0.005	0.0106	50	0.55	0.47
0.2066	−0.01	0.025	1032	0.0045	0.0064	42	0.92	0.76
0.2066	−0.01	0.025	1031	0.0022	0.0021	6	2.73	1.9
0.1516	−0.01	0.025	1145	0.005	0.0147	70	0.4	0.34
0.1516	−0.01	0.025	1032	0.0054	0.0066	66	0.89	0.78
0.1516	−0.01	0.025	1031	−0.0017	0.0042	12	1.4	0.66
0.1266	−0.01	0.02	1145	−0.0077	0.0006	3	8.66	1.35
0.1266	−0.01	0.02	1031	−0.003	0.0047	68	1.06	0.5
0.0966	−0.01	0.02	1145	0.001	0.0017	3	2.89	2.12
0.0966	−0.01	0.02	1031	0.0003	0.007	68	0.71	0.49
0.0866	−0.01	0.02	1145	−0.0063	0.0012	3	4.33	1.06
0.0866	−0.01	0.02	1031	−0.0008	0.0048	68	1.04	0.64
0.0766	−0.01	0.02	1145	−0.0012	0.0015	6	3.4	2
0.0766	−0.01	0.02	1031	0.0001	0.0047	136	1.07	0.72

database is a researched to return a more appropriate or informative data set. In the case of this research, our goal is to minimize the burden on the user and reduce the number of reformulations. A broad search formulation is made first. A refinement of the search is made only if the broad assessment does not return a usable result. The system described below was developed to facilitate broad and narrow searches as well as rapid refinement. The methods were implemented using Microsoft Visual Basic and Microsoft Excel to automatically generate the graphs.

4.1. Formulation of query

As pointed out in Section 3.5, current data access methods require the designer to either select a single run or set of runs that share some characteristics. In current systems, the user must repeatedly query the database until a reasonable size number of runs are found. To solve this problem, we have implemented an interactive query method that enables a user to quickly down select from a group of runs. The query is formulated using a set of slider bars to allow the user to quickly change the range process parameters. This method is similar to the AlphaSlider proposed by Ahlberg and Shneiderman (1994*a*, 1994*b*).

Table 5. *Incorrectly aggregated data*

Code	Bias	Std Dev	*Cp*	*Cpk*
M4.1.1 P7.1.1 F3.2	0.0037	0.0089	0.65	0.51

Figure 4 shows the prototype user interface. The user first selects the feature type (upper left corner). In addition, they specify the upper and lower limit of their target tolerance and the acceptable *Cpk* value. When the feature is selected, all relevant records are found and put in the data grid at the bottom of the screen. In addition, the data plot is automatically generated (Fig. 7) and the marker color and shape legend is built in the upper grid (the method for assigning color and shape is described below). Once the feature type is selected, the range of nominal, upper and lower limit values are automatically generated. The user can adjust the range of values and the data set and graph is automatically regenerated.

4.2. Visualization of result

As shown in Section 3.5, the current methods for presenting data are awkward. Yang-Pelaez (1999) and a number of other authors have proved that appropriate visualization methods can greatly improve a user ability to quickly traverse and understand large engineering data sets. The goal of this research was to demonstrate the usability and feasibility of using the visualization to improve the search process.

4.2.1. Graph structure

The data in the process capability database is a classic multidimensional data set. This type of data is typically represented in two-dimensional scattergraphs (Shneiderman,

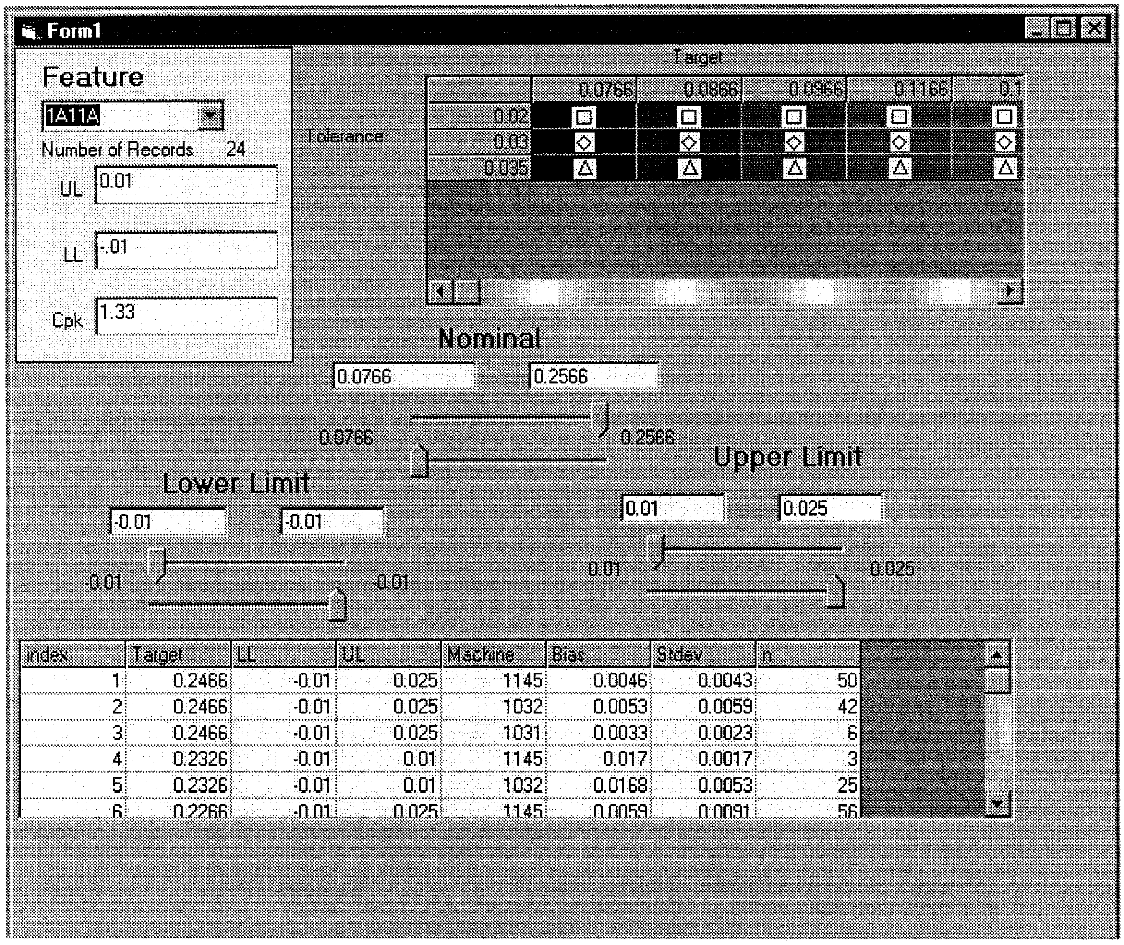


Fig. 4. User interface.

1997). The difficulty in designing visualization systems for multidimensional data is often in selecting what to dimensions should be used for the scattergraphs. However, in this case, the importance of simultaneously encoding the process bias and standard deviation is obvious as the end quality of the product is a coupled function of both.

The graph (Fig. 5) shows the data points from Table 4. Each run is plotted as a single point. The bias for each run is plotted on the X-axis and the standard deviation on the Y-axis. The acceptability limits for $Cpk = 1.33$ and $Cpk = 1.00$ are shown as lines on the graph for a tolerance band of $[-0.01,+0.02]$. The top line represents $Cpk = 1.0$ and the bottom line, $Cpk = 1.33$. Any run falling below both lines will have a Cpk of greater than 1.33; any falling above both lines will have a Cpk of less than 1.0.

4.2.2. Error bars

The standard deviations and biases are calculated from the data. Because each run has a different number of points, it is important to communicate to the designer the uncertainty associated with the sample standard deviation and bias.

A confidence interval is an interval of plausible values for the parameter being estimated. The upper and lower 95% confidence interval for the bias $[b_{\min}, b_{\max}]$ is

$$b_{\min} = b + \frac{1.96\sigma}{\sqrt{n}} \quad \text{and} \quad b_{\max} = b - \frac{1.96\sigma}{\sqrt{n}}. \tag{4}$$

The 95% confidence interval for the standard deviation $[\sigma_{\min}, \sigma_{\max}]$ is more complicated to calculate:

$$\sigma_{\min} = \sqrt{\frac{(n-1)\sigma^2}{\dfrac{\chi^2_{\alpha}}{2,n-1}}} \quad \text{and} \quad \sigma_{\max} = \sqrt{\frac{(n-1)\sigma^2}{\dfrac{\chi^2_{1-\alpha}}{2,n-1}}}, \tag{5}$$

where $\chi^2_{\alpha,\nu}$ is the chi-squared distribution, ν is the number of degrees of freedom $(n-1)$, and α is the area under the Chi-squared curve to the right of $\chi^2_{\alpha,\nu}$. For a 95% confidence interval, $\alpha/2$ is 0.025 and $(1-\alpha/2)$ is 0.975. Error bars are added to Figure 5 to communicate the confidence intervals. This allows the user to disregard points that have a high degree of uncertainty (i.e., those with longer error bars).

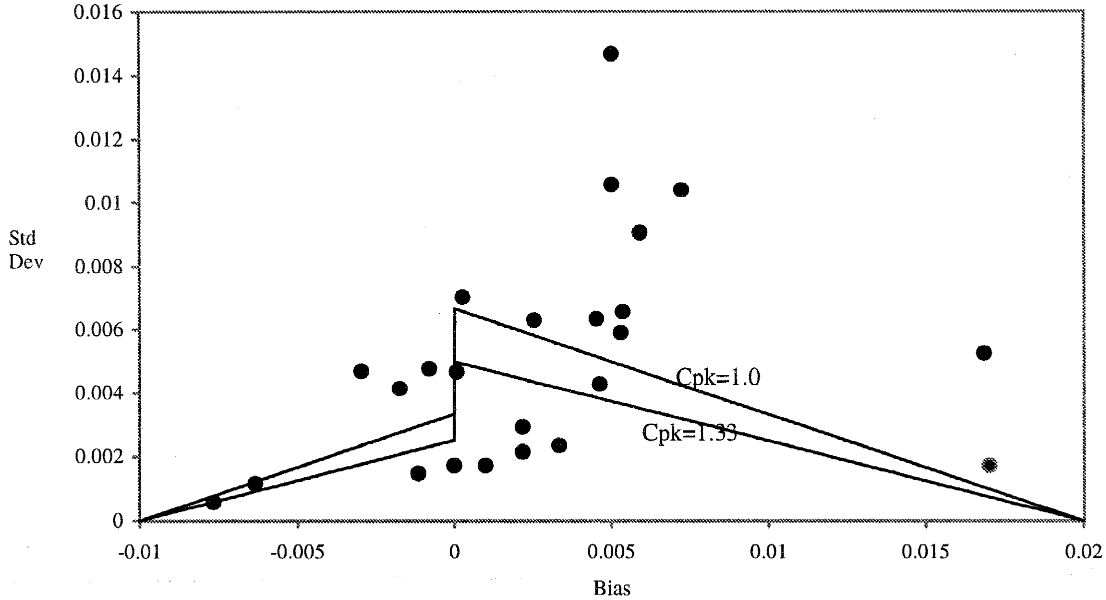


Fig. 5. *Cpk* graph.

4.2.3. Encoding

The graph in Figure 6 helps the user to assess the range of *Cpk* values. However, it is also necessary for the user to quickly assess what points are relevant if some fall below and some fall above the acceptability limit. The difficulty is in identifying what factors are relevant and how they impact capability. Relevance is a function of what process char-

Table 6. *Machine encoding*

Machine	Y	x_j
1031	0.0045	1
1032	0.0077	3
1045	0.0051	2

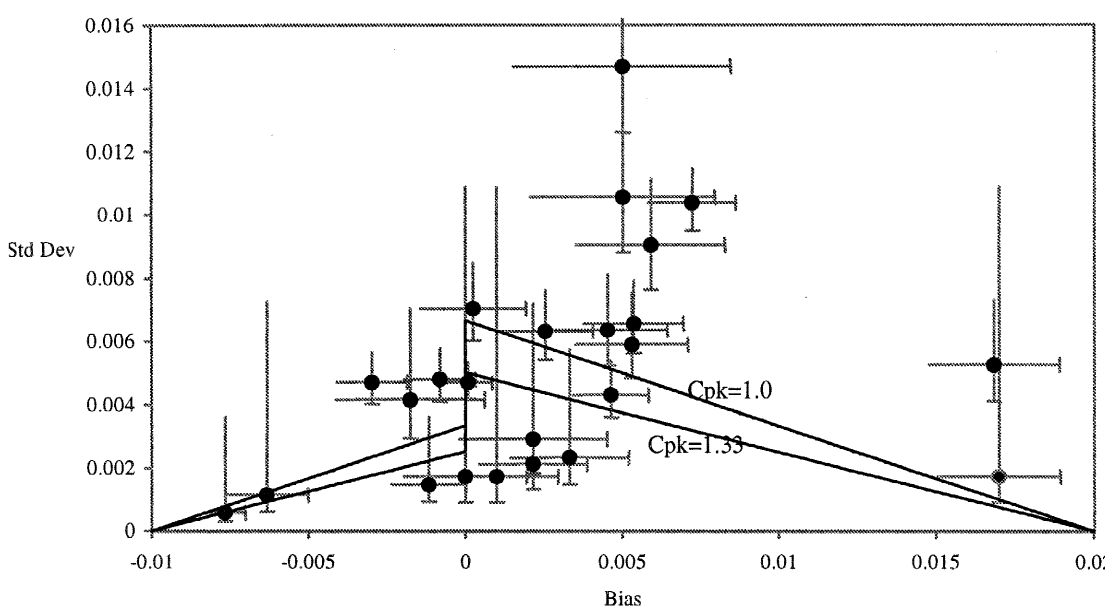


Fig. 6. *Cpk* graph with error bars.

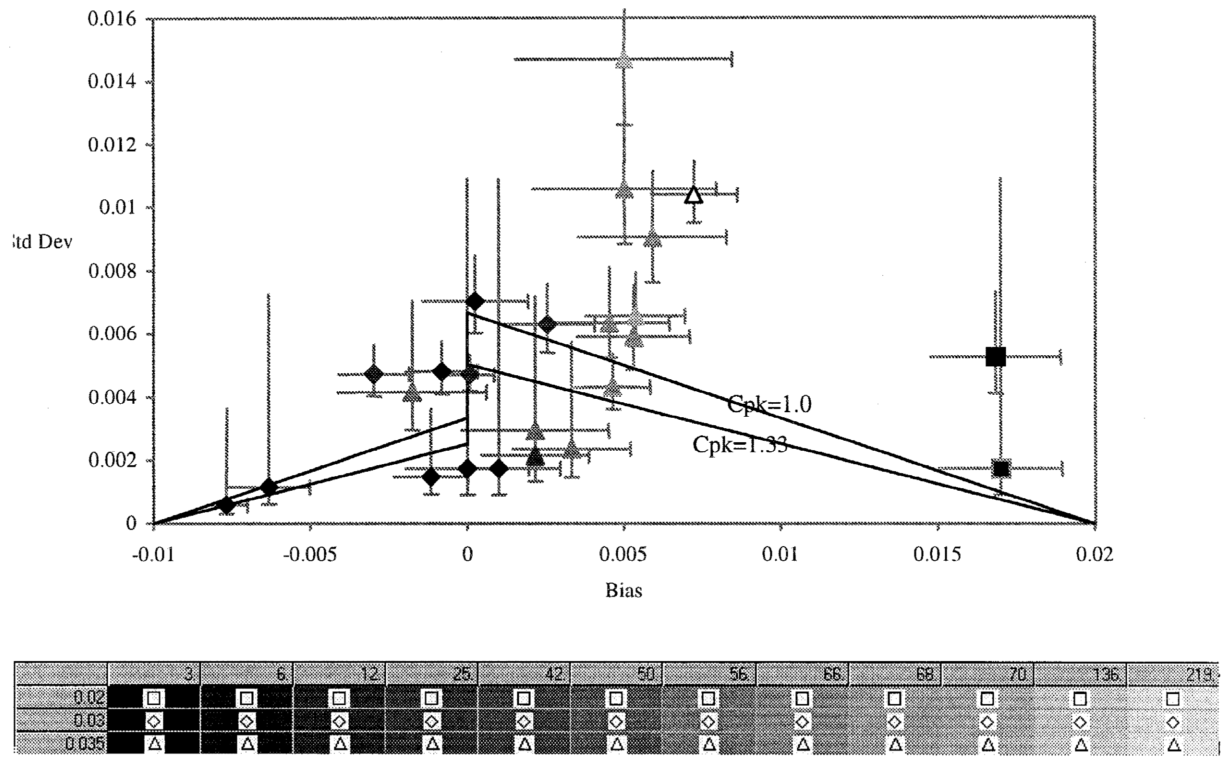


Fig. 7. Encoding with number and tolerance.

acteristics a new design shares with the surrogate run. To facilitate finding patterns, the process characteristics are encoded visually.

In a two-dimensional scatter graph, there are limited number of process characteristics that can be encoded using point color, shape, orientation, and size (Tucker, 1997). Initial tests indicated that the graph became too complex when more than two parameters were encoded. Based on user feedback, point color and shape were selected as the encoding characteristics. However, more than two process characteristics typically influence the standard deviation and bias. The simplest approach is to allow the user to select the encoding scheme. This can be time consuming for the user because they must test multiple encodings until a pattern emerges. The user must be helped to determine the most useful refinement parameters. To help the designer locate trends; the two most significant contributors are automatically identified and encoded.

The most relevant factors are automatically encoded using a multivariate linear regression analysis to identify the most significant factors. To do this, a linear model is built

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_n x_n + \epsilon \tag{6}$$

$$Y = \sqrt{\sigma^2 + b^2} \tag{7}$$

Table 7. *C_j values*

Process characteristics	C_j
Tolerance	0.0078
Machine	0.003
N	0.009
Target	0.007

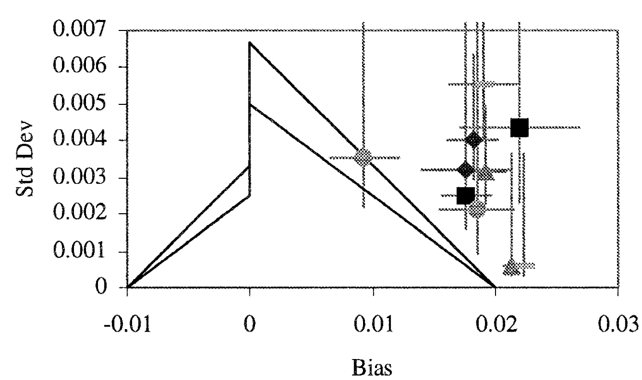


Fig. 8. All runs outside specifications.

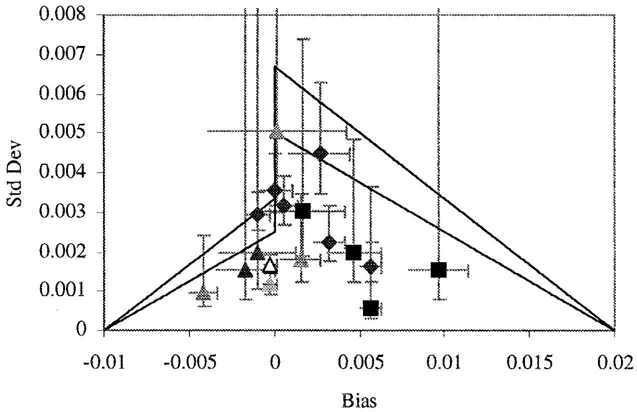


Fig. 9. All runs inside specifications.

Y is the measure of variability (based on the Taguchi loss function) and x_j is the jth process characteristic. The values of β_j are calculated using a linear regression analysis.

Some process factors may have discrete values (i.e., machine and operator). To facilitate the regression analysis, they are assigned integer values 0,1,2... based on their relative Y values. For example, Table 6 shows the average mean for the three machines: 1031, 1032, and 1045. They are assigned x values, respectively of 1, 3, and 2.

From β_j, the relative contribution, C_j, of each characteristic is

$$C_j = |\beta_j|(x_{j(\max)} - x_{j(\min)}). \tag{8}$$

C_j quantifies the relative influence variation in each term has on the overall variation. The values of C_j are rank-ordered and the largest two are encoded. Table 7 shows the C_j values for the example. In this case, the number of parts in the run is the biggest contributor and the tolerance is second.

Figure 7 shows the resultant encoding of the data in Table 4. In this case, the size of the run is encoded with color and the tolerance with shape.

4.3. Refinement

A combination of visualization and interactive querying allows a user to quickly determine the feasibility of a tolerance and determine what process characteristics are most relevant. Ideally, the answer should be found with minimum effort. In this case, we want to minimize the work required to create the original formulation and subsequent refinements.

Using these tools, the user first requests data for all runs for the feature code of interest without specifying any other

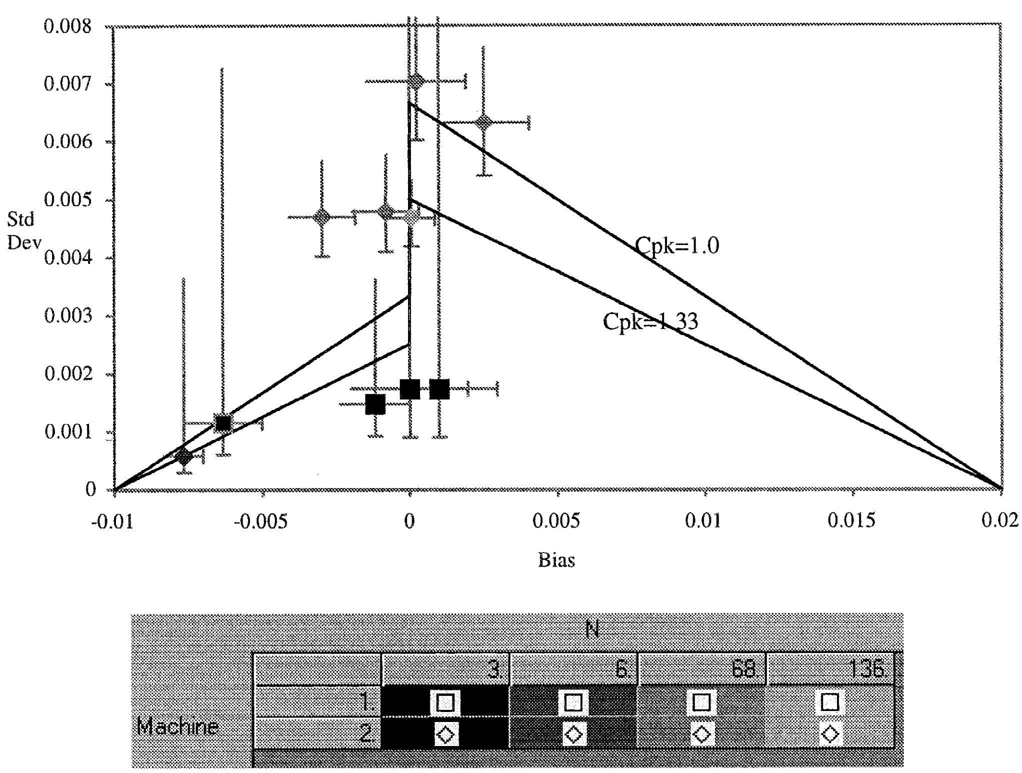


Fig. 10. Refined search.

process characteristics. There are three possible outcomes of this search. First, the process is definitely capable; all surrogate data points indicate the process will produce a feature with a $Cpk > 1.33$. Second, the process is definitely not capable; all data points indicate the process will produce a feature with a $Cpk \leq 1.0$. Figures 8 and 9 show examples of both cases. Figure 8 shows the case where all surrogate points fall outside the acceptable *Cpk* limits. Figure 9 shows the case where all points fall inside the acceptable limit.

In the third case (Fig. 7), the answer is not clear: some runs are above the *Cpk* lines and some below. The run size and tolerance were found to be important contributors. It decided to refine the search to only look at the tolerance of interest. Figure 10 shows the results when the tolerances are limited to $[-0.01, 0.02]$. It is obvious from the graph that small volumes and/or Machine 1 can produce acceptable *Cpk* values. If the part needs to be made in high volumes, it would be necessary to consult with a process expert to determine if the part could be made to the tolerances required.

5. CONCLUSIONS

In conclusion, we have shown the problem of using surrogate process capability data to be complex and nontrivial. Using the visualization method and interactive queries, it is possible to quickly assess tolerances and the factors influencing process variation. Preliminary results from industry users indicated that they found the graphical interface to be significantly easier to use, more intuitive, and more informative. Future work on this topic will include searching across multiple feature codes, building cost data into the database, automatic detection of errors, and populating variation analysis with the uncertain process capability. In addition, comprehensive user tests will be performed to fine tune the interface and visualizations.

ACKNOWLEDGMENTS

The authors acknowledge the support provided by the Center for Innovation in Product Development under NSF Cooperative Agreement Number EEC-9529140, and the Lean Aerospace Initiative.

REFERENCES

Ahlberg, C., & Shneiderman, B. (1994*a*). AlphaSlider: A compact and rapid selector. *Proc. ACM CHI '94 Conf. Hum. Factors Comput. Sys.*, 365–371.

Ahlberg, C., & Shneiderman, B. (1994*b*). Visual information seeking: Tight coupling of the dynamic query filters with starfield displays. *Proc. CHI Conf.: Hum. Factors Comput. Sys.*, 365–371.

Baldwin, R.A., & Chung, M.J. (1995). Managing engineering data for complex products. *Research in Engineering Design—Theory Applications and Concurrent Engineering 7*, 215–231.

Campbell, R.I., & Bernie, M.R.N. (1996). Creating a database of rapid prototyping system capabilities. *Journal of Materials Processing Technology 61(1–2)*, 163–167.

Chase, K.W., Gao, J.S., Magleby, S.P., & Sorensen, C.D. (1996). Including geometric feature variations in tolerance analysis of mechanical assemblies. *IEEE Transactions 28(10)*, 795–807.

DeGarmo, A., Paul, E., Black, J.T., & Kohser, R.A. (1997). *Materials and Processes in Manufacturing*. Prentice Hall, Upper Saddle River, NJ.

Frey, D., Otto, K., & Wysocki, J. (1998). Evaluating process capability given multiple acceptance criteria. *Journal of Manufacturing Science and Engineering* (to appear).

Gao, J., Chase, K.W., & Magleby, S.P. (1998). Generalized 3-D tolerance analysis of mechanical assemblies with small kinematic adjustments. *IEEE Transactions 30*, 367–377.

Kazmer, D., Barkan, P., & Ishii, K. (1996). Quantifying design and manufacturing robustness through stochastic optimization techniques. *Design Automation Conf.* ASME Design Engineering Technical Conferences

Lee, S.J., Gilmore, B.J., & Ogot, M.M. (1993). Dimensional tolerance allocation of stochastic dynamic mechanical systems through performance and sensitivity analysis. *Journal of Mechanical Design, Transactions of the ASME 115(3)*, 392–402.

Lin, C.Y., Huang, W.H., Jeng, M.C., & Doong, J.L. (1997). Study of an assembly tolerance allocation model based monte carlo simulation. *Journal of Materials Processing Technology 70(1–3)*, 9–16.

Liu, C.S., Hu, S.J., & Woo, T.C. (1996). Tolerance analysis for sheet metal assemblies. *Journal of Mechanical Design, Transactions of the ASME 118(1)*, 62–67.

Lucca, A.M., Berti, K.N., & Cerveny, D.C. (1995). Statistical tolerance allocation in design utilizing historical supplier process data. *Advances in Electronic Packaging*, *10*, 229–235.

Michelena, N.F., & Agogino, A.M. (1994). Formal solution of N-type Taguchi parameter design problems with stochastic noise factors. *Journal of Mechanical Design, Transactions of the ASME 116(2)*, 501–507.

Nagler, G. (1996). *Sustaining Competitive Advantage in Product Development: A DFM Tool for Printed Circuit Assembly*. (Master's Thesis) S.M., MIT, Boston.

Naish, J.C. (1996). Process capability modeling in an integrated concurrent engineering system—The feature-oriented capability module. *Journal of Materials Processing Technology 61*, 124–129.

Parkinson, A., Sorensen, C., & Pourhassan, N. (1993). A general approach for robust optimal design. *Journal of Mechanical Design, Transactions of the ASME 115(1)*, 75–81.

Perzyk, M., & Meftah, O.K. (1998). Selection of manufacturing process in mechanical design. *Journal of Materials Processing Technology 76(1–3), 198–202*.

Shneiderman, B. (1997). *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Reading, MA.

Soyucayli, S., & Otto, K.N. (1998). Simultaneous engineering of quality through integrated modeling. *Journal of Mechanical Design, Transactions of the ASME 120(2)*, 211–220.

Srinivasan, R.S., Wood, K.L., & McAdams, D.A. (1996). Functional tolerancing: A design for manufacturing methodology. *Research in Engineering Design—Theory Applications and Concurrent Engineering 8(2)*, 99–115.

Tata, M., & Thornton, A. (1999). Process capability database usage in industry: Myth vs. reality. *Design for Manufact. Conf., ASME Design Tech. Conf.*, (Electronic Conference Proceedings).

Ting, K.-L., & Long, Y. (1996). Performance quality and tolerance sensitivity of mechanisms. *Journal of Mechanical Design, Transactions of the ASME 118(1)*, 144–150.

Tucker, A.B., ed. (1997). *The Computer Science and Engineering Handbook*. CRC Press, Boca Raton, FL.

Wang, N., & Ozsoy, T.M. (1993). Automatic generation of tolerance chains from mating relations represented in assembly models. *Journal of Mechanical Design, Transactions of the ASME 115(4)*, 757–761.

Yang-Pelaez, J.A. (1999). Metrics for the design of visual displays of information. PhD Thesis. MIT, Cambridge, MA.

Yu, J.-C., & Ishii, K. (1994). Robust design by matching the design with manufacturing variation patterns. *Design Theory and Methodology, ASME Design Eng. Technical Conf.*, pp. 7–14.

Zemel, M.I., & Otto, K.N. (1996). Use of injection molding simulation to assess critical dimensions and assign tolerances. *Design for Manufact. Conf., ASME Design Tech. Conf.*

Zhang, C.C., & Wang, H.P. (1998). Robust design of assembly and machining tolerance allocations. *IEEE Transactions 30(1)*, 17–29.

Anna Thornton received her Ph.D. from Cambridge University in 1993, and joined the Mechanical Engineering faculty at M.I.T. January 1994 to June 1995. Professor Thornton and her students conduct research in the area of Variation risk management: the systematic identification, assessment, and mitigation of variation during product design and production. She specifically focuses on developing modeling tools to help in resource allocation and strategy decisions for variation control in design and production.

Melissa Tata received her M.S. from the Massachusetts Institute of Technology in June of 1998, and her B.S. in Mechanical Engineering from RPI. She is currently working for Dell Corporation.