The Effective Use of Process Capability Databases for Design

by

Melissa M. Tata

B.S. Mechanical Engineering
Rensselaer Polytechnic Institute, 1997

Submitted to the Department of Mechanical Engineering
In Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1999

Author ______________________________
Department of Mechanical Engineering
May 7, 1999

Certified by ______________________________
Anna Thornton, Thesis Supervisor
Assistant Professor of Mechanical Engineering

Accepted by ______________________________
Ain Sonin, Thesis Reader, Chairman of the Graduate Committee
Department of Mechanical Engineering

DISCLAIMER OF QUALITY

(Pages 195-196 are missing from the Archives copy of this thesis. This is the most complete version available.

The Effective Use of Process Capability Databases for Design

by

Melissa M. Tata

Submitted to the Department of Mechanical Engineering
on May 7, 1999 in partial fulfillment of the requirements for the
Degree of Master of Science in Mechanical Engineering

Abstract

Process capability databases have been developed at most manufacturing companies to enable process monitoring and feedback to design. The academic literature on the design tools of variation simulation analysis, robust design, and tolerance allocation assumes the existence and usability of process capability data. However, it was found that industry is far from this idea. A questionnaire was circulated to numerous industries. It revealed the current and desired uses for both internal and supplier process capability databases. This survey also showed that, although design should use process capability data to improve their designs, this is not the case.

The survey identified several barriers that prevent design from using process capability data. These hindrances are both organizational and technical. The organizational issues include the need for both better communication between functional groups and a company-wide vision of process capability usage. There are two technical issues. First, there is significant uncertainty in the process capability data. The uncertainty arises from multiple point values, outlier data, surrogate data, and aggregate data. However, this uncertainty is not communicated to the process capability database user. Second, the database interfaces are not design-friendly because the hierarchies are inconsistent, infeasible indexes are listed, and the data is displayed as an average point value.

The technical barriers to design PCD usage are addressed by providing methods to exclude outlier data, combine similar runs, group similar samples for aggregate data, determine surrogate data for unpopulated indexes, quantify uncertainty, and develop a consistent database hierarchy. A prototype software system was developed to demonstrate how existing process capability data could be presented to designers in a way that encourages its use. This was done using several methods. First, it presents the data graphically. Second, it presents the data with a confidence level in order to quantify the uncertainty. Third, it plots all the data for the parameters chosen such that several runs are depicted on one graph.

Thesis Supervisor: Anna Thornton
Assistant Professor, Mechanical Engineering

Acknowledgements

I would first like to thank my advisor, Anna Thornton, for her constant guidance, motivation, and support. She will always be a role model for me and a source of inspiration. I would like to acknowledge the Lean Aerospace Initiative for funding a portion of my research.

I want to thank my friends and family for their patience with me as I devoted all my attention to my studies and for their perpetual encouragement. I especially appreciate the support of Mom, Dad, Jessica, Tabitha, Samuel, Stephanie, Frannie, Melissa, and Jackie. I also appreciate the sharing of knowledge and friendship I have developed with all of the students in the Variation Risk Management research group. Thanks to Corinne and Bill for helping to proofread parts of this thesis.

I would like to thank my special assistant, Anthony, for his help with computer tasks, along with keeping me sane. I would like to acknowledge Hugo for helping me find a workplace where I could grow and prosper my skills.

I appreciate the help of all of those individuals who completed my industry questionnaire and/or who had discussions with me about the applicability of my research to their company. I would also like to thank the individuals who worked with my on my case study: Chris, Doug, Mario, Dave, and Michelle.

Table of Contents

List of Figures

List of Tables

Acronyms

C_p = process capability index

C_{pk} = process capability

DOE = Design of experiments

Gage R & R = gage resolution and repeatability

KC = Key Characteristic

LSL = lower specification limit

PCAR = Process Capability Acquisition Request

PCD = process capability data

PCDB = process capability data base

SPC = Statistical process control

SQL = structured query language

USL = upper specification limit

VR Coordinator = Variation Reduction Coordinator

VSA = Variation Simulation Analysis

Glossary

- **Acceptance plan** ="An acceptance plan is the overall scheme for either accepting or rejecting a lot based on information gained from samples. The acceptance plan identifies both the size and the type of samples and the criteria to be used to either accept or reject the lot" (Gaither, p. 748)
- **Aggregate data** = data provided when either all three parameters of material, feature, and process are not known or when all the details of one or more of these three parameters are not known. This data is provided as the set of samples that fulfill the details that are known
- **Control charts** = "Control charts are used to monitor the output of a process by sampling, by measuring selected quality characteristics, by plotting the sample data on the chart, and then by making decisions about the performance of the process" (DeGarmo, Black, Kohser, p. 327)
- **Confidence interval** = "an interval of plausible values for the parameter being estimated" (Devore, 1987). There is a probability percentage associated with the confidence interval, which is 95% in this thesis
- **Control parameters** = variables that differ between runs in a sample or points in a run. May be machine, operator, tool, etc.
- **Gage** = reliability of the measurement tools used
- **Histogram** = "shows the raw data and the desired value, along with the upper and lower specification limits" (DeGarmo, Black, & Kohser, p. 319)
- **Index** = set of choices for each parameter detailing data desired. The index is the label for PCD in the PCDB. Parameters included in the index vary by company, but this thesis assumes the indexes include material, feature, and process.
- **Key Characteristic (KC)** = designation on a part drawing "used to indicate where excess variation will most significantly affect product quality and what product features and tolerances require special attention from manufacturing" (Lee and Thornton, 1996)
- **Lower specification limit** = difference between lower tolerance and target
- **Lower tolerance** = minimum value for a dimension specified on a drawing
- **Mean shift** = difference between the average value and the target value for a run or sample

- **Moving average** = mean shift
- **Outliers** = data points that are substantially different from the rest of the data and thereby deviate greatly from the average value
- **Parameters** = feature, material, and process. These are values that need to be selected in the PCDB in order to obtain PCD
- **Process Capability Acquisition Request (PCAR)** = formal request for PCD
- **PCODE** = encompasses material, feature, and process. The large aerospace company uses PCODEs as indexes in their PCDB
- **Process capability** = "Process capability is a product process's ability to produce products within the desired expectations of customers" (Gaither, p. 713). "It is an indicator of what the process has done and can be expected to continue to do" (Eastman Kodak, 1995)
- **Process capability data (PCD)** = the expected and obtained standard deviations and mean shifts for a feature produced by a particular process and made of a particular material
- **Process capability database (PCDB)** = includes target and actual tolerances for particular process, material, and feature combinations
- **Process capability index (PCI or C_p)** = "PCI is useful for determining if a production process has the ability to produce products within the desired expectations of customers" (Gaither, p. 713)
- **Robust design** = process of designing a product such that it is not adversely affected by variation even though all sources of variation have not been eliminated
- **Run** = composed of points with the same index, target dimension, specification limits, machine, date, and operator
- **Sample** = composed of runs with the same value for index, target dimension, and specification limits, but with varying values for date, operator, and machine
- **Set** = composed of samples
- **Specification limits** = values for the tolerance provided with a dimension on a drawing
- **Statistical process control (SPC)** = "The use of control charts is often referred to as statistical process control (SPC)" (Gaither, p. 740). SPC is "used to ensure the ongoing quality of the manufacturing process" (Batchelor et. al. 1996)
- **Surrogate data** = data that is similar to the data for an unpopulated index

- **Target** = value for a dimension specified on a drawing
- **Tolerance** = maximum value that dimension can deviate from the specified value on the drawing. If the part is manufactured to a dimension that is greater than the tolerance plus the dimension or less than the tolerance minus the dimension, it will not be accepted
- **Uncertainty** = unsureness about the exact value. Uncertainty can be expressed as a range of possible values with some confidence interval. There are a variety of uncertainties in PCDBs including surrogate data, multiple data sets, aggregate data and small data sets
- **Upper specification limit** = difference between upper tolerance and target
- **Upper tolerance** = maximum value for a dimension specified on a drawing
- **Variation** = deviation from nominal
- **Variation Simulation Analysis (VSA)** = 3-D modeling package used to simulate the effect of variation (Ertan, 1998)
- **Z value** = quantitative method to determine if two runs (samples) are similar with 95% confidence

1 Introduction

1.1 Definition

A process capability database (PCDB) includes target and obtained tolerances for particular process, material, and feature combinations. Process capability data (PCD) is defined as the expected and obtained standard deviations and mean shifts for a feature produced by a particular process and made of a particular material. This PCD is labeled with various items including:

> "(a) Product data, such as part shape, dimensions, and specifications
> (b) Data management attributes, such as owner, revision level, and part number
> (c) Production data, such as the manufacturing processes involved in making parts and products
> (d) Operational data, such as scheduling, lot sizes, and assembly requirements
> (e) Resources data, such as capital, machines, equipment, tooling, and personnel, and their capabilities" (Kalpakjian, p. 1176).

PCD allows for an understanding of the capability of machines, tools, and operators to manufacture a particular feature of a particular dimension using a specific process. By investigating legacy PCD for similar parts/features, designers can better determine what tolerances to specify on their drawings based on capability. This assures that the tolerances are obtainable and that the design is manufacturable. The term process capability can also be used to describe geometric characteristics that a process can create, but this thesis focuses on the prior definition.

1.2 Motivation

Design for manufacturing (DFM) has received a lot of attention recently; however, design for manufacturing variation (DFMV) has been overlooked. DFMV is needed to ensure that designs conform to existing manufacturing capability. Process capability data (PCD) is needed for DFMV in the areas of robust design, optimal tolerance allocation, and variation simulation analysis. Much of the research on improving and predicting quality is premised on the existence of process capability data. However, no research discusses how to deliver process capability data to the designers in a form that they can use.

Variation reduction in manufacturing has provided benefits to many companies. For example, a number of articles in the public press have described the benefit General Electric and AlliedSignal have accrued from implementing Six Sigma methods. However, most organizations realize that they can improve the cost and quality of their products even more dramatically by improving the design of their product (rather than waiting until production to reduce variation). Ideally, designers would use PCD to allocate tolerances based on variation in similar past products.

When the process capability databases (PCDBs) were developed, the intent was for design to use PCD for optimization and product cost minimization, but this ideal situation has not been realized. Many process capability databases contain information on statistical process control, which tests to see if variation is random (chance variation) or is due to assignable causes (Fowlkes & Creveling, p. 11 and Kalpakjian, p. 1076). By using this variation information from the PCDBs, designers can more appropriately specify product tolerances because they know how much random variation to expect and how to eliminate other causes of variation.

Process capability databases (PCDBs) have been developed in many industries and are being used by the manufacturing community to monitor quality; however, they are not being effectively utilized by design for variation reduction. There are two types of issues that prevent PCD from being used by design: organization and technical. The organizational issues include the need for better communication between manufacturing and design, a company-wide vision of process capability usage, and trust between suppliers and customers.

There are two technical issues. First, there is significant uncertainty in the process capability data and it is not quantified. The uncertainty arises from multiple point values, outlier data, surrogate data, and aggregate data. Second, the database interfaces are not design-friendly because the hierarchies are inconsistent, infeasible indexes are listed, and the data is displayed as an average point value.

These technical barriers can be divided into how the PCD is presented and what items need to be added to or improved in the existing PCDB systems. There are four barriers that pertain to the presentation of the PCD. First, the process capability data is typically provided as an average point value for a series of runs with no measure of uncertainty. Second, the data is presented numerically rather than graphically and usually only one run can be displayed at a time. The data is displayed as the average value of all the runs rather than as each individual run. Third, the user interfaces and indexing schemes make it difficult for designers to obtain data because there is no consistent PCDB structure. Fourth, the user interfaces allow the user to choose infeasible indexes.

There are three barriers that pertain to items that need to be added to or improved in existing PCDBs. First, the PCDBs contain no methods to obtain alternative data when the particular index one is looking for is unpopulated. Second, there is no method to display aggregate data when designers do not know all the details of the material, feature, and process that they intend to use. Third, there is no method to eliminate outlier data or to combine runs.

1.3 Literature review

A literature review of topics related to process capability databases, variation modeling, robust design, computer integrated manufacturing, and tolerance allocation was conducted.

1.3.1 Using PCD to relate old and new products

Don Clausing discusses how rework develops from the lack of information that is available during the design stages of a product (Clausing, 1998). Oftentimes, process capability data is amongst this information that is missing from product development. Clausing also writes about the problem of "insufficient consideration given to the relationship between the product that is now being designed and other products...". Process capability databases contain information on products that have already been manufactured. If designers were trained and encouraged to use capability data on older part designs when designing new, similar parts, this would certainly speed up the design process because there is usually some reusability between products. Using this data also might eliminate some rework because designers would be developing tolerances based on their actual machine and process capabilities rather than simply on manufacturing expert knowledge.

1.3.2 Need for PCD in product delivery process

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. Similarly, Perzyk and Meftah (1998) suggest that designers should have devices to aid in selecting materials and manufacturing processes. Several articles specifically address the problem of using process capability in design for electronic systems (Lucca *et al.* 1995). Nagler (1996) proposes a design for manufacturing (DFM) tool that can be used to predict manufacturing yield earlier in product development cycle so that this information can be fed back to design. The tool was used to obtain "quantitative impacts of alternative design choices on manufacturing processes and process outcome based on historical data....".

Several authors have directly addressed some problems with process capability databases. However, most process capability database articles address characterizing the part types and geometries a process can produce, rather than standard deviations and mean shifts. 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.

1.3.3 PCD usage in tolerance allocation

Setting tolerances to match process capability and reflect design intent is the subject of significant literature (Liu *et al.* 1996; Srinivasan *et al.* 1996; Gao *et al.* 1998). A tolerance is defined as the permissible variation of a dimension in engineering drawings or designs (ANSI

Y14. 5M 1994). When tolerances are incorrectly set, rework, cost, and/or failure in service increase (Parkinson *et al.* 1993; Chase *et al.* 1996). Tolerances should be optimized to reduce mechanical errors (Lee *et al.* 1993; Lin *et al.* 1997; Zhang and Ben Wang 1998), minimize assembly problems (Ting and Long 1996), and improve product performance (Michelena and Agogino 1994; Wang and Ozsoy 1993).

Designers should use PCD to determine what tolerance values to put on their drawings. DeGarmo *et al.* (1997) stress the importance of the dimensions and tolerances that designers specify for a part. If the tolerances are too tight, "expensive and unnecessary operations result" and if tolerances are too loose or are indefinite the part may not function properly because some of its important requirements may be overlooked. Without access to PCD, designers don't understand the implications of the tolerances that they specify: "Where designers require tighter tolerances than the standard they must find out how this can be achieved, what secondary processes/process development is needed and what special control action is necessary to give the required level of capability" (Batchelor *et al.* 1996).

Ulrich and Eppinger (1995) stress the importance of designers having access to and understanding of PCD so that they comprehend the cost factors associated with the tolerances they specify.

> "A designer may specify dimensions with excessively tight tolerances without understanding the difficulty of achieving such accuracy in production. Sometimes these costly part features are not even necessary for the component's intended function; they arise out of lack of knowledge. It is often possible to redesign the part to achieve the same performance while avoiding costly manufacturing steps; however, to do this the design engineer needs to know what types of operations are difficult in production and what drives their costs." (Ulrich and Eppinger, p. 191)

DeGarmo *et al.* (1997) explain the correlation between design and production, which validates the need for PCD usage in design: "Design details are directly related to the processing that will be used, making the processing easy, difficult, or impossible and affecting the cost and/or quality" (DeGarmo, Black, & Kohser, p. 1171). PCD is also needed by designers for specifying Key Characteristics, which are designations on a part drawing "used to indicate where excess variation will most significantly affect product quality and what product features and tolerances require special attention from manufacturing" (Lee and Thornton, 1996). Designers need to be able to identify the processes that are at greatest risk for not meeting the specified tolerances. "The designer needs to understand when required tolerances are pushing the process to the limit and to specify where capability should be measured and validated" (Batchelor *et al.* 1996). Tolerances should be allocated in a manner that maximizes the robustness of the design to variation. To ensure manufacturability of their designs, designers must understand the process capability for each feature when they are specifying tolerances.

> "Any idea that designers can put tolerances on designs without consideration of the manufacturing process to be used is untenable. The designer needs to know, or else be able to predict, the capability of the process used to produce the design and to ensure the necessary tolerance limits are sufficiently wide to avoid manufacturing defects" (Batchelor *et al.* 1996).

1.3.4 Use of PCD

There are two steps to making a product more robust: predict the end quality of the design and then optimize the design. Predicting final product quality requires both a variation model and process capability data. The *variation model* takes the part and process variation as inputs, models how variation propagates through the system, and predicts the final product quality. Several tools are typically used to accomplish this: Variation Simulation Analysis (VSA), Design of Experiments (DOE) (Phadke 1989), and process modeling (Frey *et al.* 1998). The model must be populated with *process capability data* (PCD). Without accurate process capability data, it is not possible to predict the end quality of designs or to improve product robustness.

1.3.5 Academic literature assuming usage of PCD by design

A number of articles on robust design, computer integrated manufacturing, tolerance optimization, and variation modeling implicitly state the importance of process capability. The articles published in the *Journal of Mechanical Design*, *Journal of Materials Processing Technology*, *Journal of Manufacturing Science and Engineering*, *Research in Engineering Design - Theory Applications and Concurrent Engineering*, and *IIE Transactions* between 1994 and 1999 were analyzed. Twenty-eight articles in these five journals assume the existence of PCD and require it as an input to the models and tools described in the articles. A number of articles propose models to predict and optimize end product quality (Parkinson 1995; Chen and Chung 1996; Thornton 1998). Other articles describe methods to optimize product robustness (Parkinson, Sorensen et al. 1993; Andersson 1994).

1.3.6 Summary

Various articles discuss the need for PCD in product development to ensure that designers specify tolerances based on the process capability of similar old part designs. The academic literature on robust design, tolerance allocation, and variation modeling also assumes that PCD is available and used by design as input to the tools and models. The need for PCD usage by design has been identified; however, no research has been published on delivering PCD to design in a format that encourages its use.

1.4 Thesis objective

Most of the academic literature on predictive modeling and robust design assumes the existence of complete and accurate data about process capability. However, through results of surveys from several companies, this thesis demonstrates that this assumption is more myth than reality. Although companies have created process capability databases (PCDBs), the data is not being utilized by design. The PCDB studied in this thesis is amongst the state-of-the-art of the industries surveyed.

This thesis was motivated by complaints from industry about the lack of PCD usage by design. The thesis is divided into two sections. First, the current status of PCDBs and their use by industry is detailed. The barriers preventing design from more fully utilizing PCD are presented. These results are based on a survey, which was distributed to several manufacturing companies. Second, one company's PCDB was investigated in depth to determine potential improvements to increase its utility for design.

There are two classes of problem that prevent design usage of PCD: organizational and technical. The survey results revealed the organizational problems. The basis of the organizational problem is the lack of communication between manufacturing and design, trust between suppliers and customers, and a common PCDB for the entire enterprise for internal and supplier parts.

The survey identified the general technical problems of poor database user interfaces and the designer's lack of trust in PCD. However, the technical problems were primarily revealed through an in-depth analysis of one large aerospace company. The basis of the technical problems is the lack of presentation and quantification of uncertainty in the PCD and the lack of a consistent database classification scheme. Designers need PCD to be presented graphically as a range of possible values with some confidence interval rather than as an average point value. This is necessary to quantify the uncertainty of the data. Designers need to see all the runs of data rather than just the average of all the data. Designers need a consistent classification scheme for the PCDB so that they can easily determine surrogate data for unpopulated indexes and aggregate data when they don't know all the specifics for their part/feature.

According to the PCDB survey results, design needs the right usage, the right database structure, the right data, and the right management support. Potential solutions for the first two of these are presented in this thesis. Obtaining the right data and the right management support are organizational issues.

1.5 Outline

This thesis is divided into eight chapters. This chapter provides a background and overview of both related work and the work of this thesis. Chapter 2 explains the survey of various manufacturing companies which was conducted to determine how PCD is being used and the barriers preventing design from fully utilizing PCD. Chapter 2 details the desired PCDB state for both internal and supplier databases. It also explains how communication between functions, trust between suppliers and customers, and the development of database commonality across the enterprise can be used to improve the usage of the PCDBs by design by eliminating the organizational barriers.

A thorough investigation of one large aerospace company's PCDB was performed to validate the survey results and to determine the technical barriers to design usage of PCD. Chapter 3 details the background of current industry PCDBs. It provides a description of PCDB indexes, PCDB classification schemes, PCD progression, and PCD access. To demonstrate the complexity of PCD usage, Chapter 3 details the progression of data from manufacturing to quality and finally to design.

Chapter 4 describes the uses, needs and the ideal state of PCDBs at the large aerospace company. It also provides the framework for the steps that a designer would need to take to obtain PCD and the features that need to be added to PCDBs to allow this. These steps are further detailed in Chapters 5, 6 and 7.

Chapters 5, 6, and 7 are focused on how to improve the presentation of existing PCD. The basis of the improvement is to quantify the uncertainty of the PCD, to graphically display the PCD, and to develop a consistent database hierarchy. The proposed PCDB improvements have been implemented in a prototype software system, which demonstrates how PCD can be provided in a format that is understandable and amenable to design.

Chapter 5 first provides the theory for how PCD uncertainty can be quantified. Representing PCD as a point value is inadequate; therefore, confidence intervals should be added to the data. PCD is generally expressed as a series of runs or a series of samples. For these series, Chapter 4 presents examples of how they should be plotted, how outlier data can be excluded, and how the data can be combined or grouped. Chapter 5 also discusses how aggregate data can be presented to designers when they don't know all the details for the parameters of their part/feature. Finally, Chapter 5 describes the prototype software system that was developed to implement the plotting of the series of data with a confidence interval. Examples of the features of this software are presented using some PCD provided by the large aerospace company in Chapter 6.

Chapter 7 describes the hierarchy of the large aerospace company and proposes ways to make it more consistent. This chapter also proposes some methods for determining surrogate data for unpopulated indexes, which was identified as a major detriment to design usage of PCD in the survey. Chapter 8 provides conclusions resulting from this work and outlines several issues for future work.

2 Process Capability Database Surveys

2.1 Introduction

A survey of a variety of design and manufacturing companies was circulated to determine both the state-of-the-art in process capability databases (PCDBs) and the barriers preventing design from fully utilizing process capability data (PCD). Two key organizational barriers were identified for internal PCDBs: lack of a company-wide vision for PCD usage and poor communication between manufacturing and design. Supplier PCDBs have the additional barriers of lack of trust between suppliers and customers and time lag for data entry. Management support, training, database population, and common systems were identified as potential solutions to the identified barriers.

To better understand the current state of usage, as well as to understand why PCDBs are not being utilized by design, a survey was circulated to several major design and manufacturing firms. Forty-three people responded from twenty-five companies/divisions. When several divisions of one company responded, the results were generally averaged together and presented as results for the company; however, for company divisions that used to be separate companies and that make unique products, the results were kept separate. When multiple people from the same division in the company responded, their responses were averaged. The results presented were from the averaged responses from the twenty-five unique companies/divisions. Nonetheless, respondents from the same companies typically had consistent responses.

The organizations surveyed included copier, propulsion, automotive, military and aerospace product development and manufacturing firms. The survey was sent directly to the people who work with PCDBs, who helped develop PCDBs, and/or who are experts on robust design. Respondents included statistical consultants, mechanical engineering managers, design engineers, and manufacturing and quality engineers.

The survey was divided into two parts. The first part investigated the use and development of internal databases and the second, databases of supplier capability. The survey contained questions requiring both numerical and textual responses, both of which are detailed in this chapter. The quotes in this chapter came from the textual responses. A follow-up survey was generated to obtain more details for some of the responses from the first questionnaire. Twenty-one of the original forty-four respondents, representing 15 companies, completed the second questionnaire. Appendix A contains all the questions that composed both versions of the questionnaire. Appendix B contains a summary of the responses from both surveys. For each question, it is indicated how many responses were received.

This chapter summarizes the desired state of PCDBs, as described by the survey respondents (Section 2.2). The remainder of the chapter focuses on the current state of PCDBs (Section 2.3). It was found that ***PCDBs are being successfully used in manufacturing to monitor processes but are not being used to improve design***. The survey identified several technical, organizational, and informational barriers to design usage of PCD (Section 2.4).

- Poor population of PCDBs.
- Data pertinent to design not available.
- Lack of management support.
- Lack of usage metrics.
- Lack of incentives for PCD use.
- Lack of PCDB commonality across enterprises.
- Lack of direct design access to PCDBs.
- No linkages to other information systems.
- Poor indexing schemes.
- Poor user interfaces.
- Out-of-date PCD.
- Design's lack of trust and understanding of data.
- Design's lack of PCDB training.

These barriers are caused by two fundamental problems: failure to communicate between design and manufacturing and a lack of a common, enterprise-wide approach to PCDB usage in the product delivery process. Potential solutions to these barriers are also proposed (Section 2.5) based on the analysis and the respondents' future improvement plans. The key to improving design's usage of PCD is giving designers the ability to get the right data quickly. A similar analysis of usage and barriers was done for supplier databases (Section 2.6). A summary of these supplier PCD barriers is given in Section 2.7. Supplier PCDBs contain PCD for parts from suppliers and may be separate from or together with a company's PCDB for internal parts. Supplier PCDBs have some additional barriers to design usage:

- Separate PCDB for supplier data.
- Confidentially of supplier PCD.
- Lack of consistency and availability of supplier PCD.

Four topics were covered by the survey: PCDB desired state (Section 2.2), current usage (Section 2.3), usage barriers (Section 2.4), and future solutions (Section 2.5). The results for the supplier databases are separate from the internal part databases because they have two unique barriers (Sections 2.6 and 2.7). The results for both internal and external databases are detailed in Section 2.8. Several of the questionnaire results are detailed in a *Design for Manufacturing* conference paper to be published in 1999 (Tata and Thornton).

2.2 Desired state

PCDBs were originally designed for use by both the manufacturing and design communities. Figure 2.1 shows the percentage of respondents who indicated that their PCDB was developed for process monitoring, design feedback, inspection, or regulatory requirements. Other development reasons included corporate metrics, dimensional management, and variation simulation analysis.

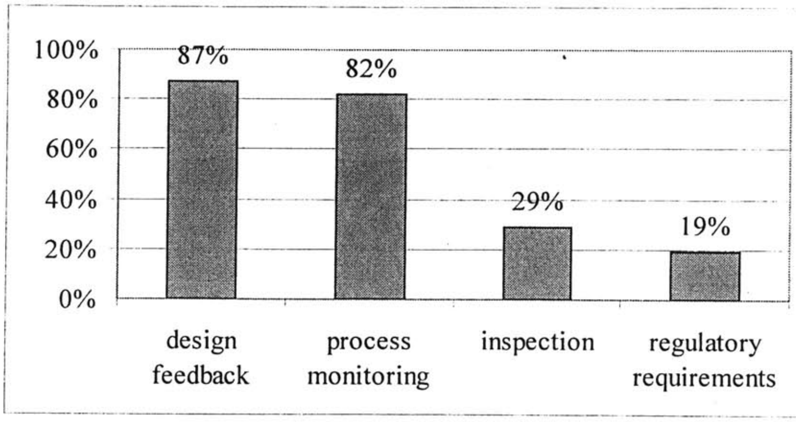


Figure 2.1: Desired PCDB usages

Eighty-two percent of the respondents indicated that they would like to use internal PCD for designing new parts with more appropriate tolerances. Respondents also identified several ways PCDBs could be used to improve quality and reduce costs in the design process: identify areas to apply robust design, specify realistic tolerances, and enable design quality verification prior to production. One company would like to "generate an exception report for characteristics that do not meet six sigma." Companies would like to use PCD to "determine the feasibility of critical parameters", to "determine the interaction between subsystems and parts", to "decrease time to manufacture a new part by using knowledge of prior parts", and to "determine optimal process(es) to make a particular part/feature".

Ideally there would be a "lessons learned database that could be accessed by any site to see best practices and problems encountered by other sites" and "data and knowledge would be transferred to the next generation of a product family for improving time-to-market." Companies would like to use PCD in design to: "design out variation when required", "establish tolerances and key characteristics for a product", "make products more producible", "make designs more robust", "simulate variation", "prioritize process improvements", and "understand the cost impact of parameter values." Companies would like to use databases to "decrease time to manufacture a new part using knowledge of prior parts."

The survey results showed that the ideal PCDB is fully populated with up-to-date and accurate data. In addition, it links directly to computer-aided design (CAD) packages and simulation software (i.e., VSA). The ideal database estimates manufacturing costs to enable design trade-off analyses. Ideally, the system would automatically "caution designers when a feature or manufacturing process is being considered that will not meet the established quality level for that particular program." Companies would like to be able to do "cost and cycle time trades vs. performance." Finally, many companies would like to see "a direct link to a drawing program to automatically flag tolerances that do not meet established quality levels."

2.3 Current state

Companies want to use PCD in design to improve product quality and producibility. Most responding companies (95%) have some type of PCDB; however, PCD is used on only ***28.9%*** (Figure 2.2) of projects and most companies (71%) use it less than thirty percent of the time. The usage level refers to the number of critical projects/subsystems where process capability was used to validate the design prior to production. Eighty-one percent of the respondents indicated that they use process capability data for design at some level.

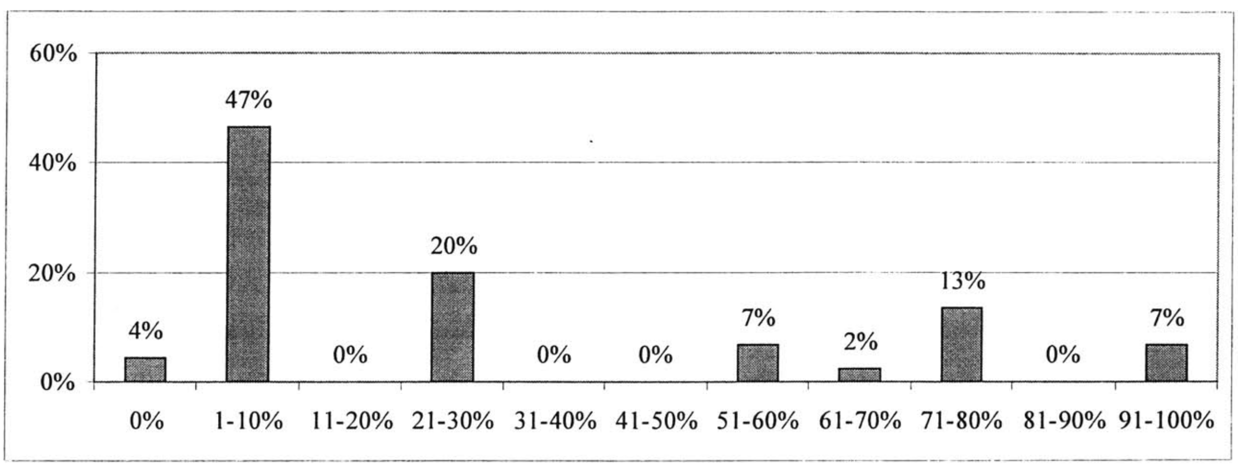


Figure 2.2: Percentage of projects employing PCD

In addition, the PCDBs are still relatively new (Figure 2.3) with an average age of 3.4 years among the respondents.

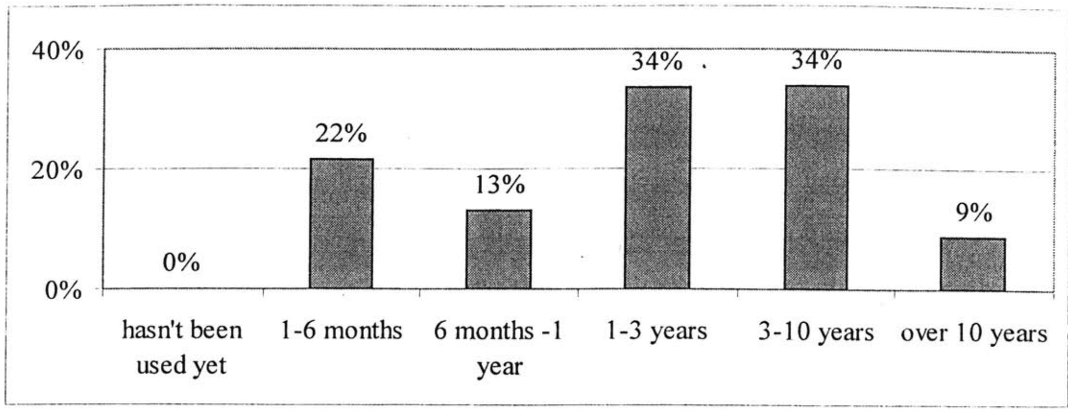


Figure 2.3: Age of PCDBs

Figure 2.4 shows how designers currently use PCD.

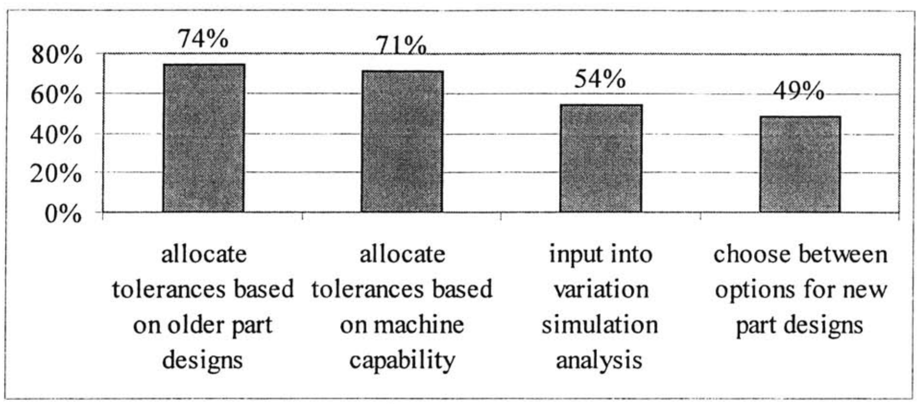


Figure 2.4: How designers use PCD

However, Figure 2.5 shows that only ***30.9%*** of the tolerances specified by designers are based on real process capability data. Instead, Figure 2.6 shows that most of the tolerances are set based on manufacturing expert knowledge. The respondents indicated that 16.7% of the time their tolerances are allocated based on variation simulation analysis, 24.3% based on robust design, and 20.6% based on guesses about capability.

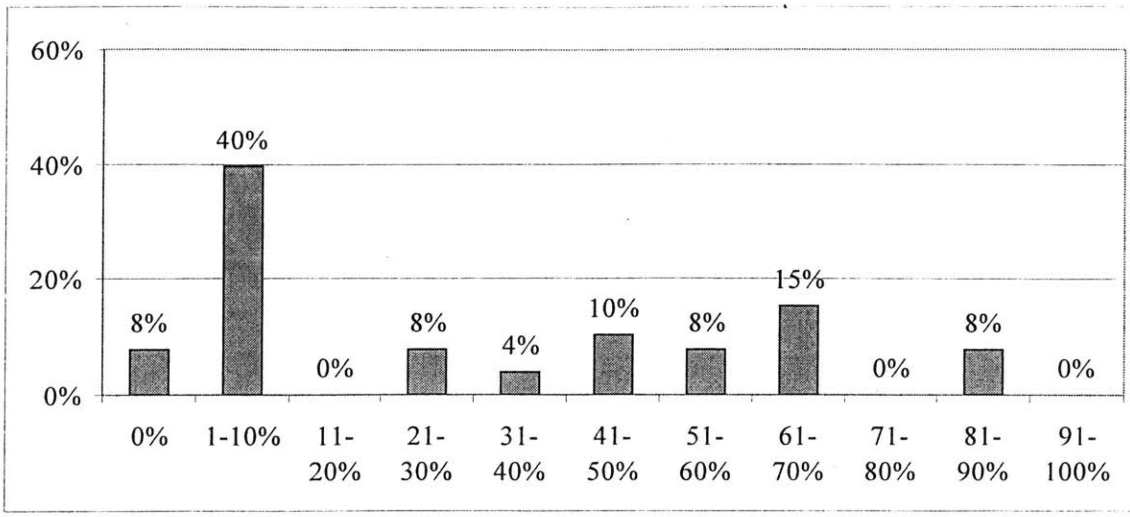


Figure 2.5: Tolerances based on real PCD

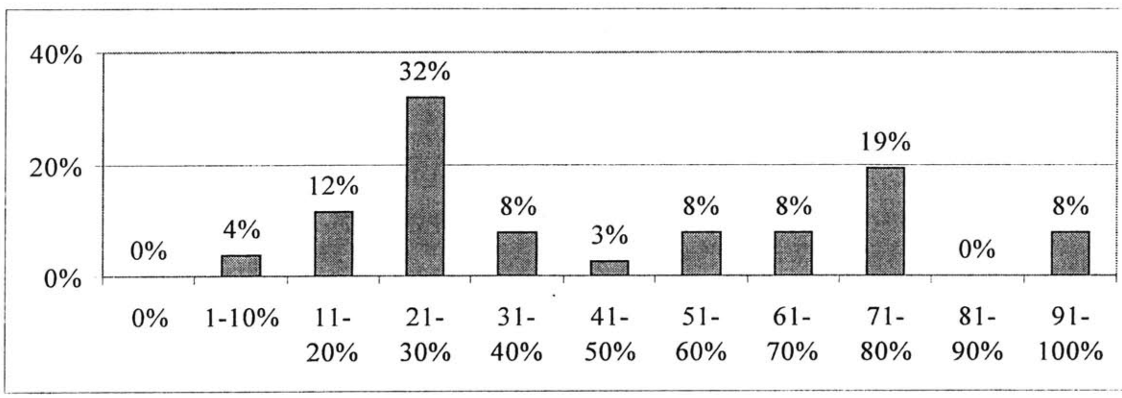


Figure 2.6: Tolerances based on manufacturing expert knowledge

2.4 Design usage barriers

Most of the survey was dedicated to identifying barriers to PCD usage by design. The most prominent barriers are poor population of databases, lack of needed data, lack of management support, and limited accessibility to PCDBs. Other obstacles include no linkages between PCDBs and other information systems, lack of usage metrics, poor user interfaces, poor PCDB indexing scheme, design's lack of trust and understanding of data, out-of-date data, no incentives to use PCD, lack of design PCDB training, and lack of database commonality across enterprises.

The following sections describe each of the barriers. The summary in Section 2.4 describes the interrelations between them.

2.4.1 Poor population of PCDBs

Most databases are not fully populated; an average of ***38.1%*** of internal parts are contained in databases (Figure 2.7). Two factors contribute to this: the databases are fairly young (Figure 2.3) and data has not been entered consistently.

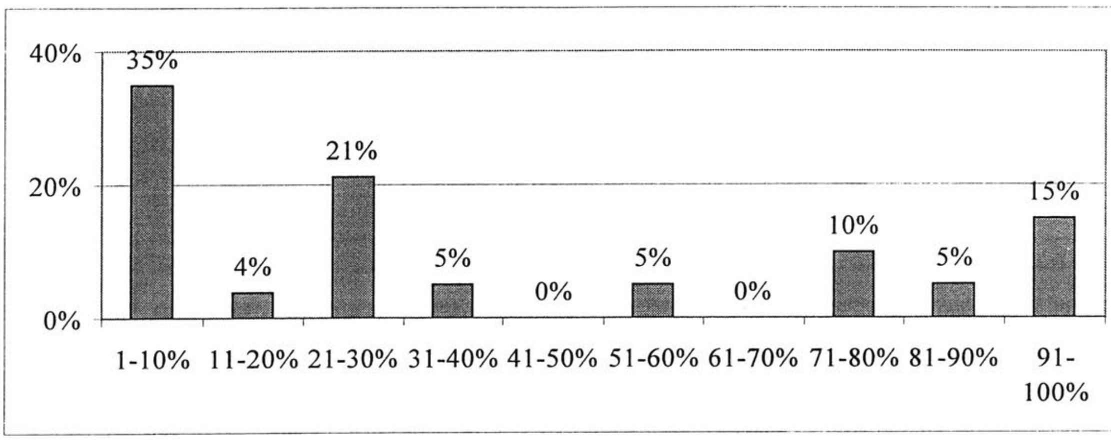


Figure 2.7: Percentage of internal parts in PCDB

There are particular types of data that are most often populated in the database; however, these vary substantially between companies. The respondent's databases are populated with data for parts: that were manufactured most recently (33%), that have automated data entry (8%), that are most expensive (8%), that have tolerances that are the most critical (13%), that contain no data already (2%), that have undergone a process improvement (25%), that are required most frequently (10%), and/or that are the newest (17%).

The lack of PCD significantly limits design's ability to verify quality. If a designer repeatedly queries the database and the required information is not available, he/she will typically stop utilizing the database.

The respondents indicated why their databases are not populated with data for all of their internally manufactured parts: Figure 2.8.

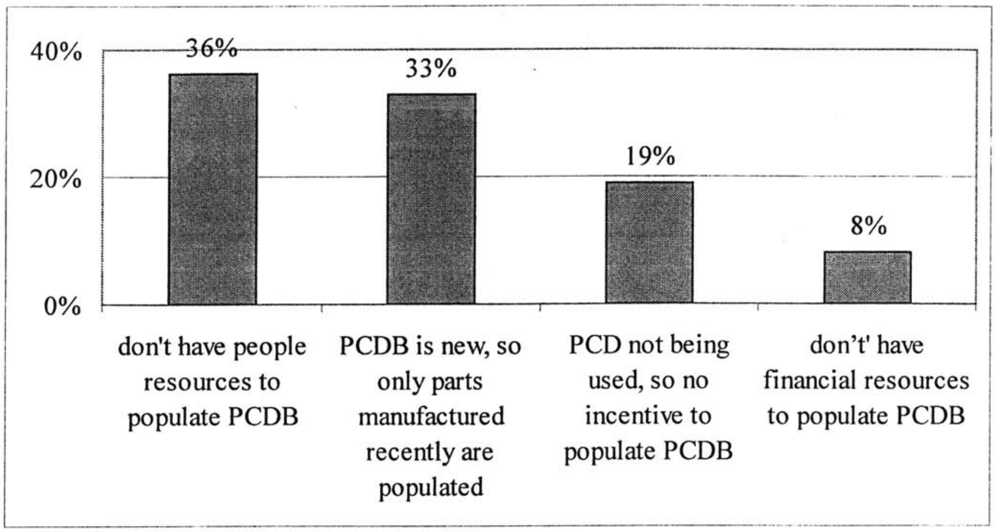


Figure 2.8: Why database not fully populated

2.4.2 Data pertinent to design not available

According to one operations excellence specialist, "data doesn't match what designers are looking for." The data used to monitor process performance and the data needed by design are often not the same. Although manufacturing collects statistical process control (SPC) data (84%), key characteristic data (54%), and part data (59%) only the key characteristic data is typically requested by design. The SPC data is used to control processes and part data is used for inspection and/or process variation monitoring. Manufacturing engineers indicated that they would be willing to collect the data specifically for design; however, designers typically have not been proactive in identifying what types of feature/process/material data they need.

Designers indicated which information they would like to see contained in the PCDBs: Figure 2.9.

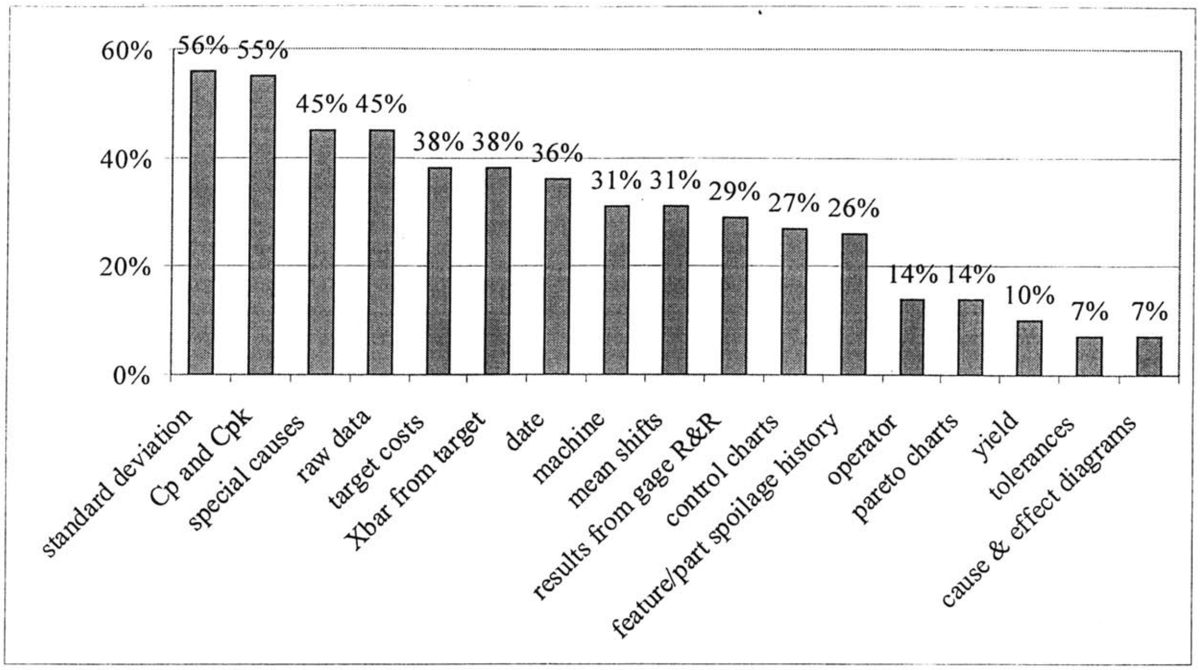


Figure 2.9: PCD design wants

2.4.3 Lack of management support

Forty-nine percent of respondents listed lack of resources as an obstacle to design PCDB usage –"it is difficult to get the PCDB prioritized high enough to get it implemented." PCDBs require significant resources including equipment, data maintenance, and training. Because PCD is not being used by design, many companies are now questioning the value of their existing investment –"managers do not have a clear understanding of why PCD is needed, nor do they understand the amount of time and effort that is required to collect and analyze the information." In the last year, many companies have withdrawn support for PCDBs. Resources arc needed to improve and maintain the PCDBs. Surprisingly, almost half of the respondents (42%) indicated that their PCDB has received increase funding during the past year. Forty-two percent of the respondents had decreased funding and 15% had constant funding.

2.4.4 Lack of usage metrics

The management support problem is aggravated by the lack of good metrics to track database usage. Sixty-three percent of the respondents do not track frequency or patterns of usage. This is due to a number of problems including lack of resources. One company monitored data usage in the past, but found that people were taking credit for obtaining the data from the PCDBs although they were not using it to improve their designs.

2.4.5 Lack of incentives for PCD use

During the design process, management is not requiring or rewarding the use of process capability data. As one manufacturing engineer pointed out "designers are not required to look at PCD as part of their design process." The lack of incentives is a barrier according to forty percent of respondents.

2.4.6 Lack of PCDB commonality across enterprises

Over seventy-eight percent of the databases are locally developed and maintained (Figure 2.9). In addition, databases within the same enterprise tend to be incompatible. A wide variety of software packages are used (ACCESS (23%), ORACLE (20%), EXCEL (20%) and QUANTUM (24%)) and the indexing schemes are not compatible. Incompatibility and dispersion of PCD was identified as a major hindrance by fifty percent of respondents.

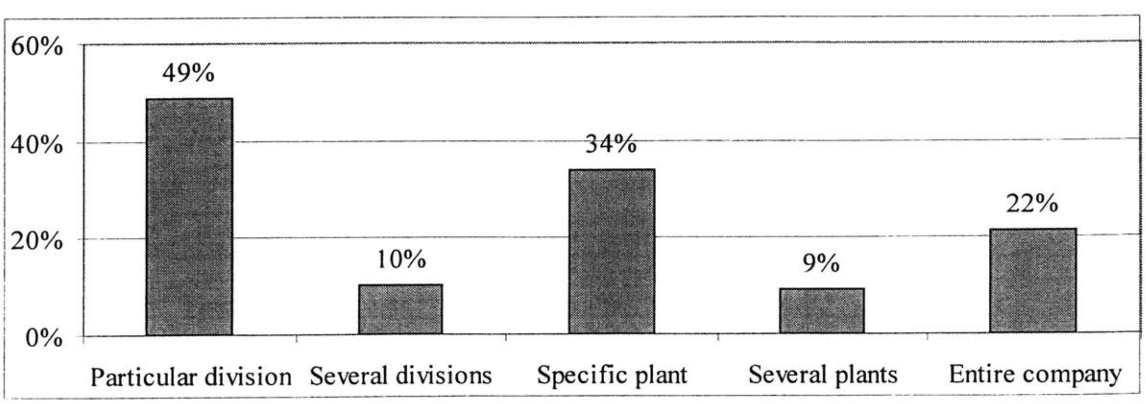


Figure 2.10: Location of PCD

Figure 2.11 shows that most companies use PCD only at their particular site/division. The respondents indicated that design uses the PCD the most followed by manufacturing and then quality.

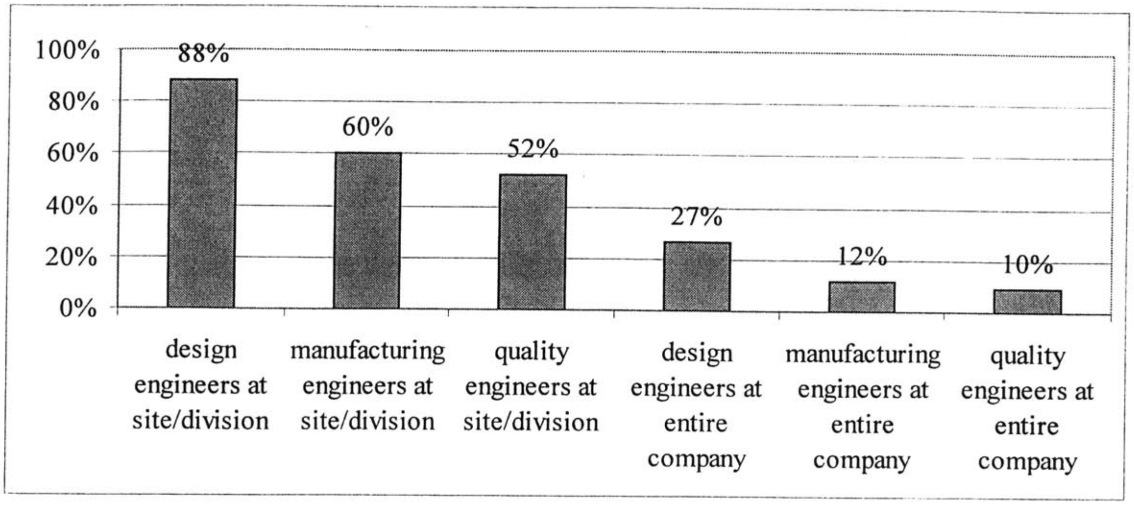


Figure 2.11: Who uses PCD

2.4.7 Lack of direct design access to PCDBs

Forty-eight percent of the respondents have PCDB access available to all company employees. The other fifty-seven percent limit access to specific groups: process engineers, product delivery teams, quality engineers, operators, design engineers, supervisors, and/or mechanics. A variety of reasons are given for limiting access (Figure 2.12).

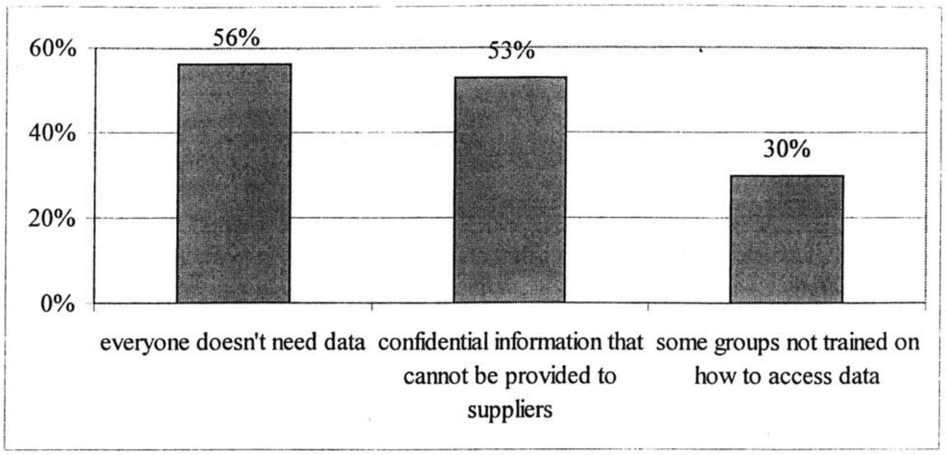


Figure 2.12: Reasons for limited access to PCDBs

Several manufacturing engineers indicated that they don't provide designers with direct access to the PCD because they don't know or trust how a designer will interpret and use the data – "the combination of the database design and the lack of education on process capabilities, lead users of the data to look for the wrong data and apply it incorrectly to the design." Without direct access, designers must submit PCD requests to the manufacturing engineers. The PCDB owners work with design to determine what data they need and then interrogate the database for them. This process tends to be very time-consuming. One operations excellence specialist notes that "designers don't have time to wait for PCD" and another that "design engineers are behind schedule and don't have time to obtain the data."

Even if designers are granted access to the PCD, data access is awkward. The data is accessed through multiple access methods: shop floor computers (50%), the intranet (48%), network servers (27%), request forms (12%), or weekly and monthly reports. Even when designers have intranet or network access to the PCDB, many do not have the software to access the data. One operations excellence specialist indicated that "access to (PCD) is available but not automatic – you need to know who to ask for it to get it."

Figure 2.13 shows that designers usually obtain PCD from manufacturing, reference manuals, or data requests. However, 39% of the responding designers have direct PCDB access.

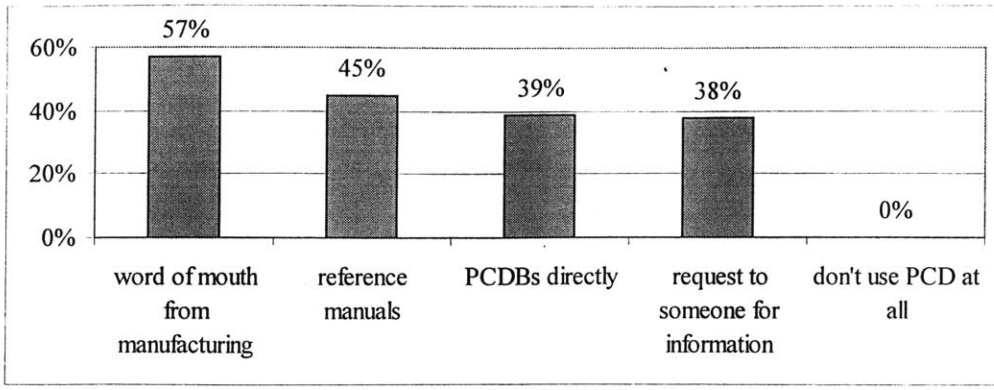


Figure 2.13: How designers obtain PCD

2.4.8 No Linkages to other information systems

Another major barrier to effective process capability data usage by design is the lack of linkages to other information/analysis systems –"CAD systems don't interface with PCDB." Figure 2.14 shows how few links companies have between their databases and other systems. Most of the linkages are pointers from the database to other systems. For example, many databases point to the part drawings but not to the specific feature. None of the companies have systems that enable designers to access PCD from modeling systems such as VSA.

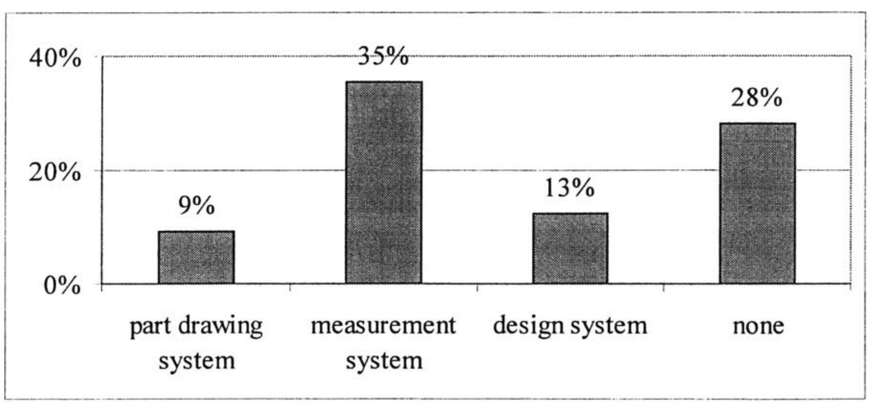


Figure 2.14: Links to external systems

Respondents were also asked which specific design programs are linked to their PCDB. Fifty percent have links to computer-aided design tools, 30% to variation simulation analysis tools, and 20% to design of experiment tools. When designers use these linkages, they almost always (80%) have to copy the data from the PCDB into these other systems rather than having a direct link (0%).

2.4.9 Poor indexing schemes

Another problem comes from the indexing schemes –"data is not being characterized properly such that it would be useful for the design community even if they wanted to use it." Designers typically want to access data by the feature, material, and process characteristics of the designs they are creating. However, "data is not indexed by query desired" because manufacturing usually indexes data by the part number or key characteristic number. In this case, searching for the appropriate surrogate process capability data requires an understanding of all of the parts in the database. Surrogate data is needed when the desired data is unpopulated in the PCDB. Thirty-two percent of the respondents identified the PCDB structure as a barrier. One said "the database can be easy for the manufacturing function to enter data and use it, but the design function cannot readily use it."

Fortunately, several companies have begun to index their databases based on material, process, and feature characteristics. Fifty-two percent of the respondents said they access the PCDB data by feature type and 72% by manufacturing process. However, the companies we have visited have multiple indexing systems at the same site and/or have not completed the process of re-indexing legacy data systems – "there is a lack of integration due to fixed mentalities or old paradigms" according to one engineer/scientist specialist.

2.4.10 Poor user interfaces

Generally the PCD is presented in numerical format and only one set of data can be viewed at a time. In many cases, the user interface requires detailed knowledge of both database query languages and the structure of the specific database. A material and manufacturing process engineer said "there is no user-friendly interface and only those that can write SQL queries can get data." One respondent said that the "software and graphics are complex and difficult to utilize". Many respondents also agree that "the PCDB software is not easy to work with" because they have "limited analysis capabilities" and "limited flexibility."

The PCD is presented in the form of either raw data (50%), control charts (37%) or histograms (37%).

2.4.11 Design's lack of trust and understanding of data

In many cases, designers don't trust the process capability data – "engineers don't know about the data, trust the data, or trust the location of the measurements." Data might not be reliable if both shifts don't collect data, if different datum schemes are used, and if different measurement systems are used (Leland, 1997). One senior manager for variability reduction indicated that "manufacturing-collected data may not always be reliable/accurate." First, the databases often don't include a measure of statistical validity including number of data points in a population or gage resolution and repeatability data. Second, special causes of variation are often not indicated. Third, the indexing schemes may not have significant resolution. As a result, the data returned for a certain process index may have significant variability. Fourth, in some cases the date is not included with the data. In the automotive industry, Leland (1997) found that "Differences between functional groups...serve to create" a "lack of trust in other's data."

There are many causes for the designer's lack of trust in the data that could be eliminated. One problem is that 19% of the time when data is added to a index that is already populated, it is simply averaged with the old data. This could be alleviated by keeping the old data separate from the new data and labeling it with its date. Then, it is still possible to average the old and new data if desired. This is currently being done by 47% of the respondents.

Process improvements and problems are also not optimally noted. Only 14% of the respondents have notes linked to particular PCD while another 21% have data for particular process, material, feature, etc. separate for each process improvement. The other respondents either have notes separate from the PCDB (48%) or don't record this improvements/problems at all (21%).

2.4.12 Design's lack of PCDB training

Designers also are often not trained on how to use PCDBs –"the data is not user-friendly to access or to interpret." One senior manager for variability reduction indicated that "design does not always know what to do with the data." Sometimes the designers don't even know that their company is collecting PCD – "designers don't know PCDB exists."

2.4.13 Out-of-date PCD

There is a time lag between when the data is generated and when it is available; however, design needs access to the most up-to-date data. The time lag results from the data being entered manually. Less than half of the respondents have the PCD entered automatically. One respondent indicated that because of "manual data entry, PCD is updated infrequently."

2.5 Summary for internal process capability databases

The barriers described in Section 2.4 are highly coupled. To better understand the relationships between the barriers to design usage of PCDBs, a cause and effect diagram was built (Figure 2.15).

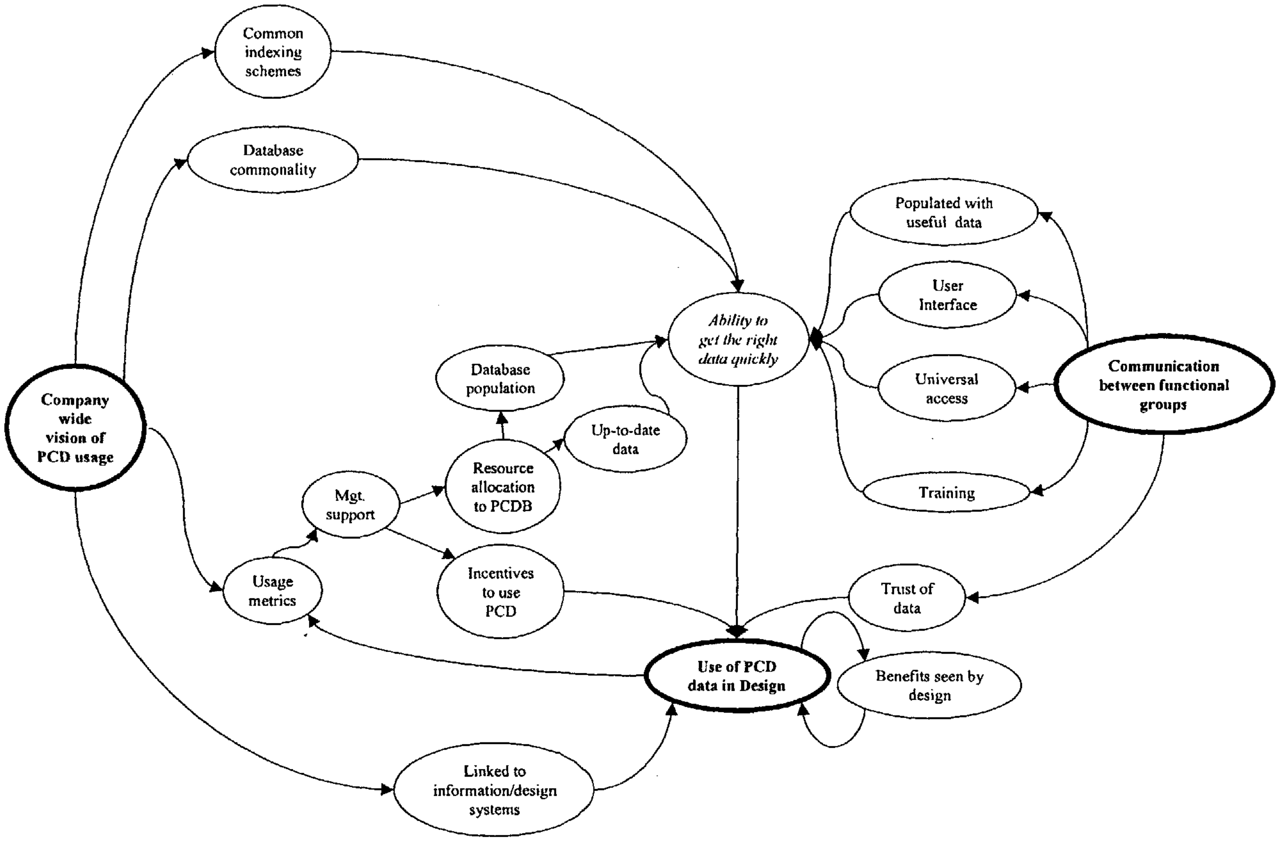


Figure 2.15: Cause and effect diagram for internal PCD design usage barriers

Figure 2.15 shows that the fundamental need of design is the ability to get the right data quickly. In order to do this, six needs must be met. First, the PCDBs must be fully populated with the type of data desired. Second, the PCD must be updated frequently. Next, designers must be trained on using the PCDB. Fourth, the PCDB must have a feature-based indexing scheme to make it easy for designs to find the information they need. Next, the database must be common across the entire company. Finally, designers must have access to the PCD for all divisions of the company.

Design's resistance to using PCD was found to be due to two root causes: a lack of a company-wide vision and plan for process capability database usage and a lack of communication between functional groups such as design and manufacturing. Designers need the PCDBs to be linked to information/design systems, need to trust the PCD, and need to be able to obtain the desired data quickly. Once they use the PCD, they will see the benefits of doing so, which will encourage them to continue using it.

2.5.1 Company-wide vision

A company-wide vision of PCD usage is needed because of the distances between when, where, and who generates and uses the data. During production, manufacturing needs to collect and maintain the correct set of data in a form that design can use and trust. Then, during new product development, design should use this data to validate their designs and to set appropriate tolerances.

It is hoped that using PCD in design will produce visible benefits. However, there are two additional barriers. First, the analysis of process capability, manufacturability, and robustness requires design to invest extra resources when resources and time are most constrained. Second, the benefits of design efforts are not accrued until the design is transferred to production.

A company-wide vision should make four improvements: implement common indexing schemes, develop database commonality across the enterprise, streamline the process by investing in linking PCDBs to other information/design systems, and implement PCDB usage metrics. The usage metrics will show that the PCD is being used by design and thereby will induce management support for the systems. Supportive management will allocate resources to the PCDBs and will provide designers with incentive for using PCD. Monetary and people resources will allow for the PCDB to become more populated with current data.

Several companies are considering developing one PCDB for their entire company to alleviate the problems of training, access, and data population. This appears to be a good solution; however, transferring legacy systems, ownership, updating duties, and maintenance are major obstacles to such an endeavor. In addition, unless improved indexing schemes are introduced, searching a monolithic database will be very cumbersome. A better idea, which some companies plan to try, is to "develop a shared server access for all sites so data can be easily accessed from any site."

2.5.2 Better communication between design and manufacturing functional groups

Although integrated product teams exist in many companies, manufacturing and design don't communicate enough about PCDBs. Manufacturing engineers have been in charge of setting up and populating the databases; therefore, they have tailored databases for process monitoring. Designers have not been active in this development; therefore, their needs have not been met. "It would be easier to establish good communication between manufacturing and design if there were more trust and understanding of the benefits that manufacturing inputs can provide and the limitations of the manufacturing process capability" (Nagler, 1996).

Better communication between design and manufacturing should make five improvements to design's use of PCD: PCDB training, universal PCDB access, user-friendly PCDB interface, database populated with the data that design needs, and trust of PCD by designers. One respondent summarizes the need for design to understand PCD: "The design community, in general, does not understand process capability and its use in the product definition process. Incentives and management support will make them want to use the data, but without proper understanding, it will be used incorrectly, which may be a bigger detriment."

With the limited resources for PCDBs, it is important that design informs manufacturing about which data is most crucial to populate PCDB with. If manufacturing focuses on inputting historical data, they may miss the current data. On the other hand, manufacturing cannot simply populate the PCDBs with the current data or else the statistical validity of the data will not be adequate because of the small sample size. "As ... manufacturing processes become more complex, yield can decrease, which makes good communication between design and manufacturing even more important" (Nagler, 1996).

2.6 Supplier databases

In today's product development organization, a company rarely produces all of the parts and sub-systems in a product. In most cases, upwards of half of the parts in a product are procured from outside suppliers. When designing a system, it is necessary to have access to both internal and supplier process capability data. Historically, parts were designed and then sent out to suppliers for price quotes. In this case, contractual obligation and piece part inspection were used to ensure compliance to the tolerance requirements. However, as suppliers become more like partners, it becomes more important to communicate process capability. Supplier PCDBs share some of the same problems as internal PCDBs; however, they also have some unique challenges.

Only about half the companies (48%) with internal part PCDBs also maintain supplier PCDBs; nonetheless, several other companies indicated that they plan to develop supplier PCDBs soon. The companies who do not maintain supplier data indicated that it's the supplier's responsibility to maintain capability data and to make it accessible on request. Two reasons were given for developing supplier PCDBs: to design better systems (50%) and to choose between suppliers (63%). Other uses for this information include: evidence of supplier process control, improved supplier processes, supplier certification, histogram qualification, appropriate design change identification, key characteristics, and datum selection. However, as stated above, most development efforts do not make use of process capability data when designing parts; nonetheless, ninety-two percent of the respondents indicated that they would like their supplier PCD to be used by design. Supplier data is usually collected by materiel/procurement groups. These groups require suppliers to report process capability data as part of contract requirements.

Figure 2.16 shows that only ***14.8%*** of the internal parts are designed/toleranced using supplier PCD.

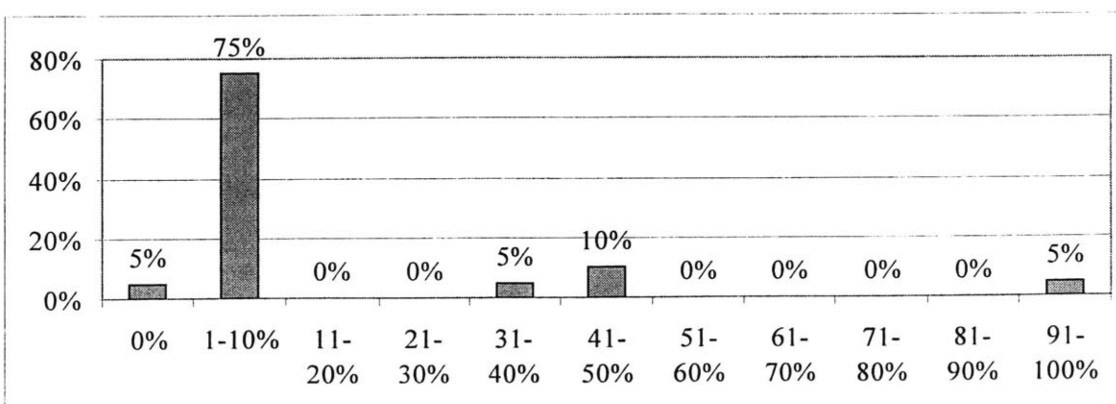


Figure 2.16: Percentage internal parts designed/toleranced using supplier PCD

2.7 *Design usage barriers for supplier data*

Design does not use PCD from suppliers for the same reasons they do not use data for internal parts: the databases are poorly populated (an average of ***43.9%*** of supplier parts are contained in PCDBs), there is a lack of management support for the systems, there is no PCDB commonality, there is a lack of direct design access, and the PCDBs have poor indexing schemes. Several of these common issues are aggravated by supplier specific issues including time lag and confidentially.

2.7.1 Poor population of supplier PCDBs

First, in many cases, suppliers don't provide data to customers. Second, suppliers typically provide the data only for the particular part the customer has ordered. The same processes are often used for multiple customers; however, the customer is only given a small percentage of the available data. Sixty percent of the respondents have PCD for less than forty percent of their supplier parts as shown in Figure 2.17.

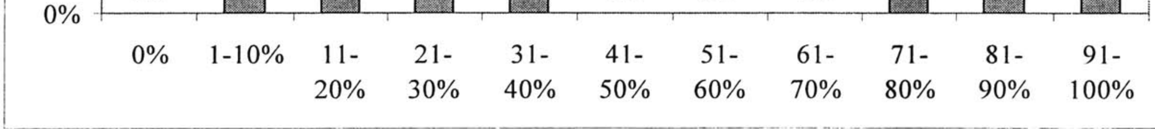


Figure 2.17: Percentage of supplier parts in PCDB

2.7.2 Lack of management support for PCDBs

One senior manager for variability reduction said that his/her "supplier management organization doesn't have the resources to manage and track the data." Another said, "implementation is stalled due to other priorities." Many companies feel that "it is the supplier's responsibility to produce and supply their customers with acceptable, defect-free products and service." One company indicated that their purchasing group wouldn't cooperate to develop a supplier PCDB.

2.7.3 Lack of PCD commonality across enterprises

Seventy percent of the companies/divisions who maintain supplier data, keep it in a separate database from the internal data. Having two separate databases makes it more difficult for designers and other employees to access the correct information.

2.7.4 Lack of direct design access to supplier PCD

Only thirty percent of the respondents provide universal internal access to supplier data. The reasons limiting PCDB access are shown in Figure 2.18. Most companies have agreements with their suppliers not to share their data with other suppliers. Most engineers access the supplier PCD through the intranet (49%), shop floor computers (19%), the network (11%)or a request form (11%).

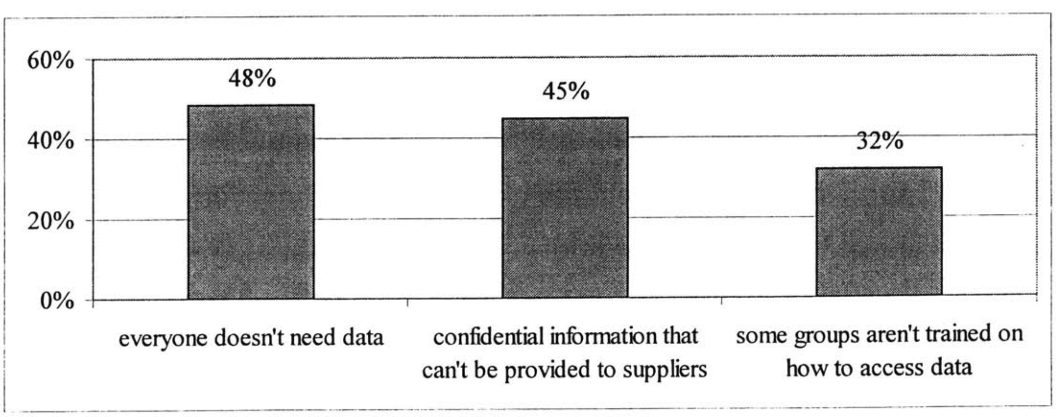


Figure 2.18: Reasons for limited access to supplier data

2.7.5 Poor PCDB indexing schemes

A wide variety of indexing schemes are used in supplier systems. The data is accessed through part number (83%), key characteristic number (35%), feature number (30%), manufacturing process (26%), feature type (26%), and machine (21%). Other indexing methods include: tooling, suppliers, team, product, and material. The proliferation of indexing schemes is aggravated by the lack of an industry standard.

2.7.6 Supplier PCD not readily available

There is a significant time lag between when PCD is generated and when it is accessible. Supplier data is often (64%) entered into the database by quality groups. The manual entry and the variety of formats lengthen the entry time. Another problem is the inconsistency of supplier data. Data arrives in a variety of formats from different suppliers. Suppliers have multiple customers each of which have unique process capability data reporting requirements. Customers have many suppliers each of whom may provide the data in a different format. Forty-five percent of the respondents receive data in a handwritten format, 41% in a process capability program, and 53% as a spreadsheet.

Other forms include: formal report submittals, qualification reports, on-site reviews, weekly and monthly reports, and histogram reports. Some companies are considering the possibility of streamlining supplier data so that all suppliers provide data in the same format. One technical advisor for process improvement indicated that "all supplier data must be transferred to a standard format", so all supplier data should be obtained in this format originally. Nonetheless, one respondent indicated that "the vast majority of sub-tier suppliers have too many different customers that would demand too wide an array of reporting. This would drive suppliers' costs well beyond any perceived value."

2.7.7 Confidentiality of supplier PCD

Suppliers are hesitant to share process capability with customers and/or designers because of two problems: confidentiality and competitiveness. The first is a risk that other suppliers will be allowed to access the data, even though most companies have "an agreement with each supplier not to share their data with other suppliers." The second problem is caused by the need stated in Section 2.6; fifty-eight percent of the respondents want to use process capability data to choose between suppliers.

2.8 Summary for supplier process capability databases

Figure 2.19 shows the cause and effect diagram for the supplier PCD design usage barriers. The causes and effects are superimposed on the previous cause and effect diagram for internal PCDBs (Figure 2.15). The supplier-specific problems stem from two root causes. First, communication between design and materiel impacts the same issues as for internal PCDBs and also directly impacts the ability to get the right data quickly (due to time lag). Second, supplier relations result in varying data requirements, different formats, and the need for confidentiality agreements. There is also a problem with the lack of commonality between PCDBs for internal and supplier parts.

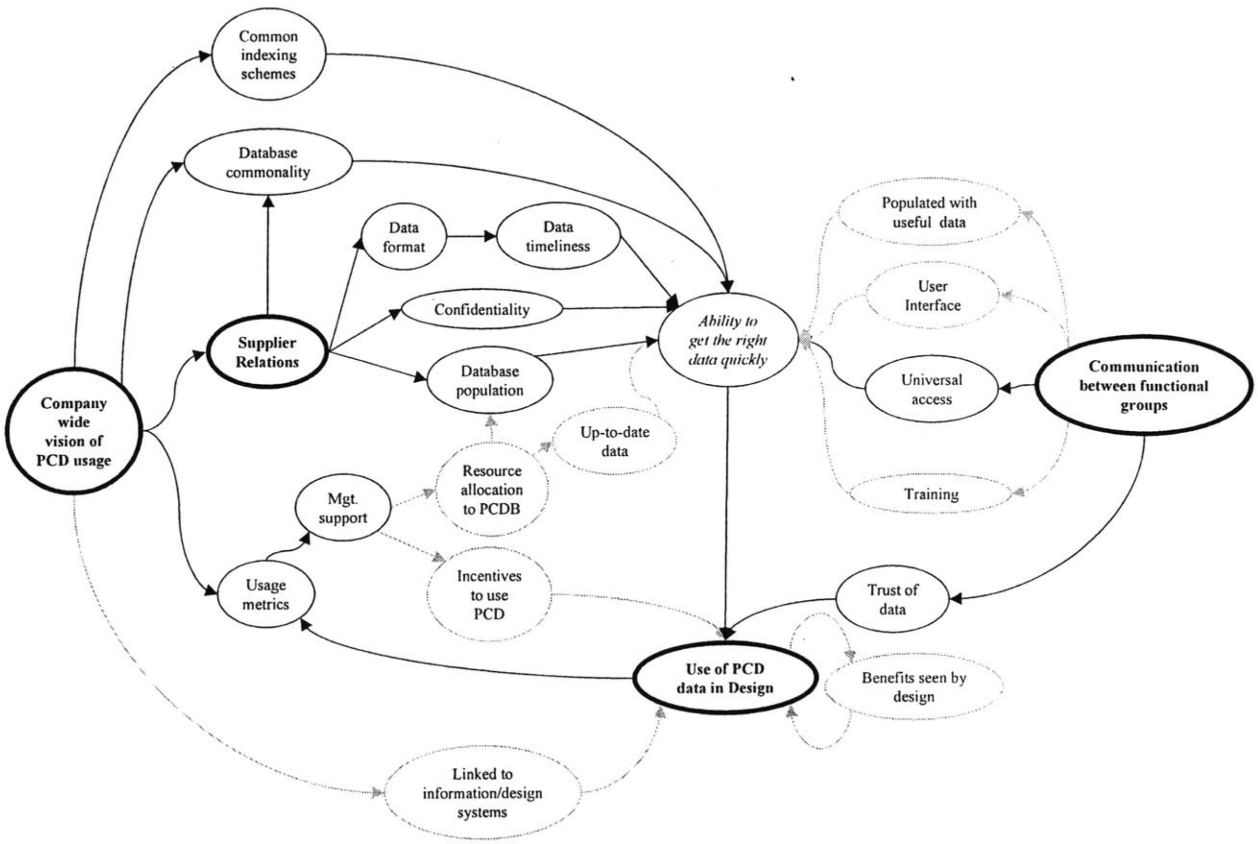


Figure 2.19: Cause and effect diagram for supplier PCD design usage barriers

Companies have many plans to improve their supplier PCDBs. The first is to increase the reporting requirements. In addition, companies plan to integrate their supplier and internal databases. Some companies plan to enable electronic transfer of supplier data directly into their database. One company plans to "allow suppliers access to data that they submitted, associated data from internal parts that mate with their parts, and assembly measurements."

One large hindrance to companies including supplier data in their process capability database seems to be an unsureness in how to prevent supplier's from seeing other supplier's or the main company's proprietary information. According to Owen, aerospace companies often won't send files "over the wire" at all because they are more concerned about maintaining their privacy from their competitors than they are about wasting time. One company, Industrial Design and Imaging (ID&I), seems to have found a solution to this. "ID&I sets up a home page on its Web site dedicated to that customer or the customer's product, complete with passwords and file encryption. Renderings, drawings, manufacturing capability studies, lists of materials, and costs are all there" (Owen, 1998).

2.9 Conclusion

Initially, PCDBs were developed for both process monitoring and design feedback. However, the goal of design feedback is not being achieved because of three reasons (Figure 2.20). First is a lack of communication between design, manufacturing, and materiel. Second and third are a lack of trust between suppliers and customers and a lack of a company-wide vision about how to utilize process capability data in the product delivery process. These are the organizational barriers to PCD usage by design.

In order to utilize the current PCDBs for design feedback, several fundamental changes must be made. First, the incentives and processes to use the data must be implemented. Without this, even the best database will not be used. Second, a company-wide strategy for the database structure must be developed. This will facilitate training and alleviate accessibility issues. Third, communication between functional groups to identify what data should be collected and how it should be presented and interpreted must be improved. Figure 2.20 shows that the barriers to PCD usage by design can be split into four main needs. The PCDB must have the right structure, the right data, the right usage, and the right management structure.

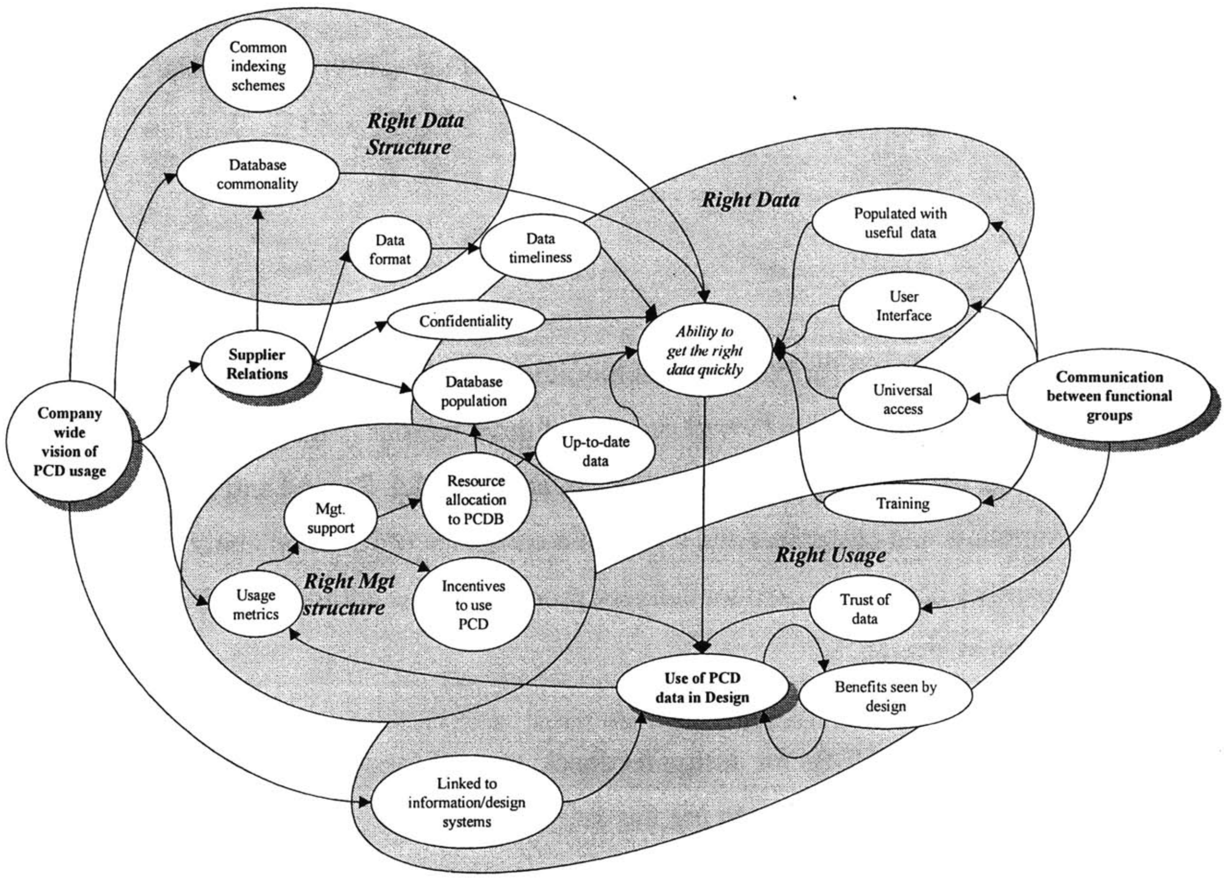


Figure 2.20: Combined cause and effect diagram for internal and supplier PCD design usage barriers

There are groups and companies that are successfully using process capability during their design process. However, these examples of success are limited and tend to be found in small pockets in the organization. The success of these projects is more often due to the strength of the design team, rather than any information or organizational tools.

The respondents were asked what parts of their PCDBs are in greatest need of improvement. These results are shown in Figure 2.21.

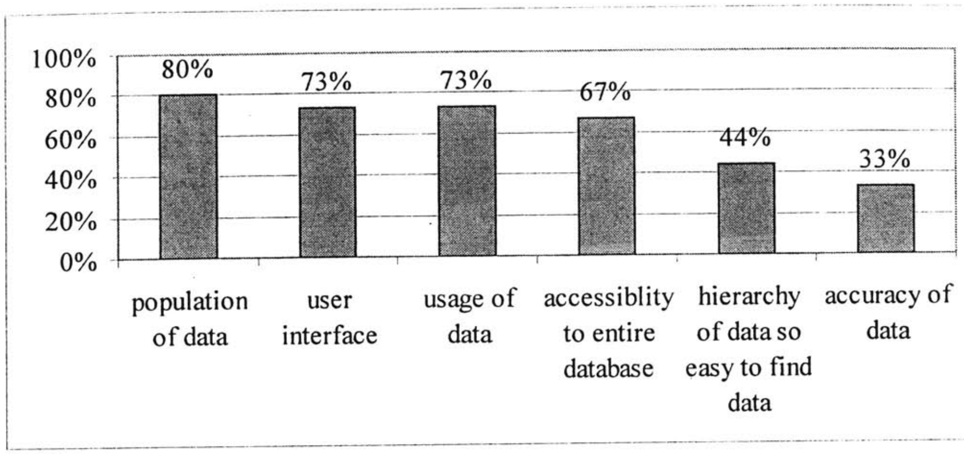


Figure 2.21: Parts of PCDB in greatest need of improvement

Design needs the right data to be fully populated in the PCDB. They need the database to have a friendly user interface. Designers need to begin to use PCD so that they can see its benefits and use it more often. Designers need universal access to the databases and they need the databases to have a hierarchy structure in which it is easy to find data. Designers also need to trust that the PCD they obtain is accurate. Overall these results show that the greatest current need for improving PCDBs is having the right data, followed by having the right structure and the right usage. Having the right management structure is secondary. The further research detailed in Chapters 5, 6, and 7 focus on the three primary issues of the right data, structure, and usage.

Finally, the responding companies/divisions indicated what incentives would prompt designers to use PCD if their PCDB was fully populated: Figure 2.22. In order to encourage design use of PCD, designers need to know the benefits of using the PCD. Seventy percent of the respondents indicated that they would be willing to participate in a case study to prove PCD usage for design. After the improvement needs identified in Figure 2.21 are made, the management structure for PCDB usage needs to be developed. Management needs to both require and provide incentives for PCD use by design. Also, the PCDB structural changes need to ensure that designers can obtain the data quickly.

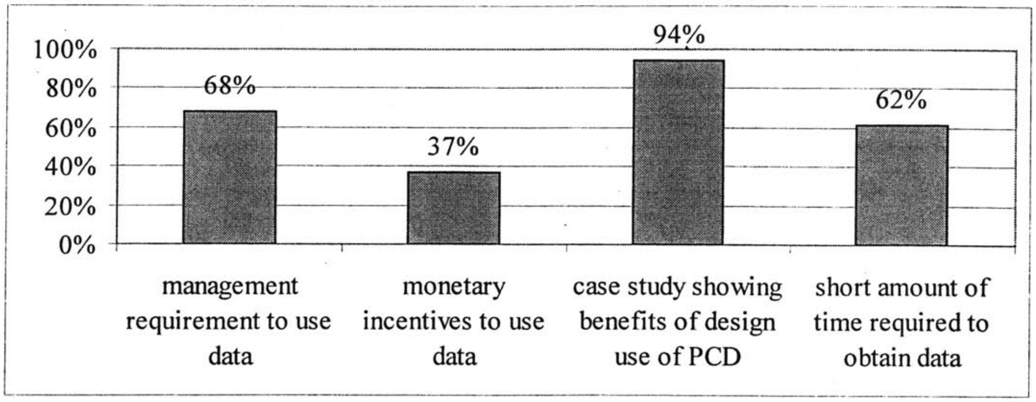


Figure 2.22: Incentives to prompt designers to use PCD

3 Background of Current Industry Databases

3.1 Introduction

This chapter provides background on the current state of process capability databases in industry. First, it describes the case study that was performed in order to determine the technical issues preventing design from using PCD. Next, it describes the indexes that are used to find the process capability data and the database classification scheme, which is composed of these indexes. Then, examples are provided to show why it is not statistically valid to provide PCD as a point value for the average of all runs for the desired index. Instead, uncertainty should be added to the data and the designer should be able to see all runs for the desired index. Finally, this chapter describes the progression of PCD and design access to PCD at the large aerospace company studied.

3.2 Case Study

The results of the surveys detailed in Chapter 2 and discussions with several companies indicated that companies have process capability databases (PCDBs) and realize their potential usefulness. Having process capability data (PCD) enables robust design, tolerance analysis, process modeling, and key characteristics. Nonetheless, companies have not been able to realize the full benefit of the PCD because it is underutilized in product development.

A case study was performed to determine how one company's process capability database could be improved to enhance its use in product development. These improvements were implemented in a prototype software system designed to enhance data validity and visualization. In order to validate the software, it was developed and evaluated in conjunction with a large aerospace company. This company has a similar process capability database to other large companies in the aerospace, automotive, and consumer products industries. The large aerospace company's database encompasses similar problems to those identified in the questionnaire results detailed in the previous chapter. However, its database population and indexing scheme are good compared to most of the companies surveyed.

In order to develop the PCDB improvements, a two step process was used. First, it was determined how the large aerospace company currently uses their PCD. This is detailed in Section 3.6. Next, it was determined how this large aerospace company and other companies want to use their PCD in design, production, and supplier management (Section 4.2). Finally, the PCDB information needs of supplier management, the customer, quality, manufacturing, and design to enable these PCD uses are detailed in Section 4.3.

3.3 PCDB Indexes

Each process capability database has some type of index to identify the data. The index represents the data label and often is composed on the material, feature, and process for the data. When manufacturing inputs the process capability data, it is given an index based on its characteristics, which might include KC number, part number, material, feature type, process, machine, operator, feature number, tool, department, fixture, etc. In order to retrieve the PCD from the PCDB, the designer must input the desired values for each of these characteristics. The large aerospace company whose database was studied uses PCODEs to index their PCDB. These PCODEs are composed of material, feature type, and manufacturing process and contain a maximum of seven numbers and letters.

The indexes of all companies consulted are usually composed of several numbers and/or letters representing each characteristic. A fictitious example follows in Figure 3.1.

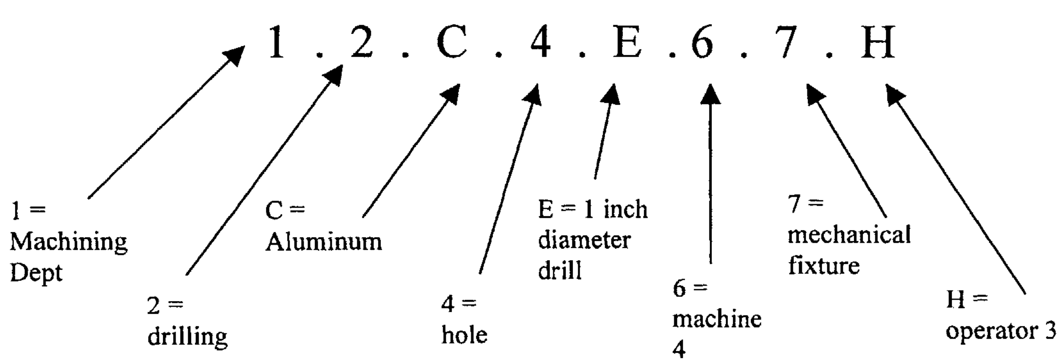


Figure 3.1: Example of an index of numbers and letters

For each populated index, there are typically several runs of data. All runs have the same material, feature, and process, but may vary by machine, operator, and /or date. The process capability data for a particular index is usually presented as a point value that represents the average of all the runs in that index.

3.4 Classification Scheme

The classification scheme of a PCDB is the way that the indexes are divided. For example, one classification scheme that the large aerospace company uses is PCODEs. Each PCODE is composed of a material, feature, process, tool, department, and sometimes a fixture type. Prior to this classification scheme, the large aerospace company had a scheme in which the index was composed of a material, a feature, or a process. In this scheme, the designer would have to choose three separate indexes in order to find the data for the material, feature, and process desired.

The advantages of the new scheme, where the material, feature, and process are all encompassed in the index, is that it is much easier and faster to find the data desired. Also, the infeasible combinations of material, feature, and process are eliminated. Only those material, feature, and processes that can be combined are given index values. The disadvantage of this classification scheme is that it has more digits in the index and more indexes. Assuming that there are 100 feature types, 100 materials, and 100 processes, there were 300 indexes in the first system where material, feature, and process were separated. However, there were about 1,000,000 possible index combinations that could be chosen with this system. In the new classification scheme with material, feature, and process combined in the index, all infeasible options would be eliminated, so there would be less than 1,000,000 indexes. Nonetheless, there would be significantly more than 300 indexes with this new scheme.

Figure 3.2 shows an example of the detail of the PCODEs in the old classification scheme for the large aerospace company. An example of each of material, feature, and process is given.

1	Material
1.1	Material Aluminum
1.1.1	6111-T6

2	Process
2.7	Hole preparation
2.7.3	Counterbore
2.7.3.2	Semi-automatic

2	Feature
2.1	Hole
2.1.1	Blind
2.1.1.1	Depth

Figure 3.2: Examples of PCODEs indexes

The PCODEs have various levels of detail. This varying level of detail allows the data to be accessed at either a general (1 or 2) or a detailed (1.1.1 or 2.1.1.1) level.

The various hierarchy problems discussed in Chapter 7 are evident in both the old and the new classification schemes of the large aerospace company. Most companies still have classification schemes where the user has to select an index for each parameter, like the old scheme used by the large aerospace company.

3.5 Run depiction

In many current databases only a point value for the average of all the runs is provided. The designer would choose the index to match the feature/part he/she is producing, then he/she would input the desired maximum standard deviation for that part. The desired standard deviation is equivalent to the tolerance that the designer would put on the drawing. In order for the tolerance to be acceptable, the standard deviation of past PCD for the desired index must be less than the desired maximum standard deviation. If only a point value for the average of all the runs is provided, then only this value can be compared to the desired standard deviation. Figures 3.3 and 3.4 show why this is unacceptable.

In Figure 3.3, the average point value for all the runs of the desired index is less than the desired maximum standard deviation; therefore, the designer would assume his/her tolerance is acceptable. The computation of standard deviation is detailed in Section 5.2. However, adding the confidence interval to the average value shows that the upper confidence interval for the average value for all the runs of the desired index is greater than the desired maximum standard deviation. This shows why adding a measure of uncertainty to the PCD is essential. The confidence interval is "an interval of plausible values for the parameter being estimated" (Devore, 1987). There is a probability percentage associated with the confidence interval and in this thesis it is 95%. For the data in Figure 3.3, the designer would need to increase the tolerance to exceed the confidence interval for the average value for the desired index.

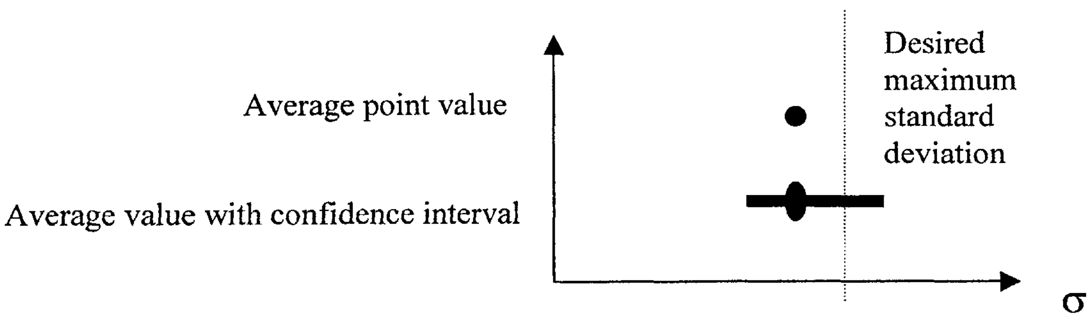
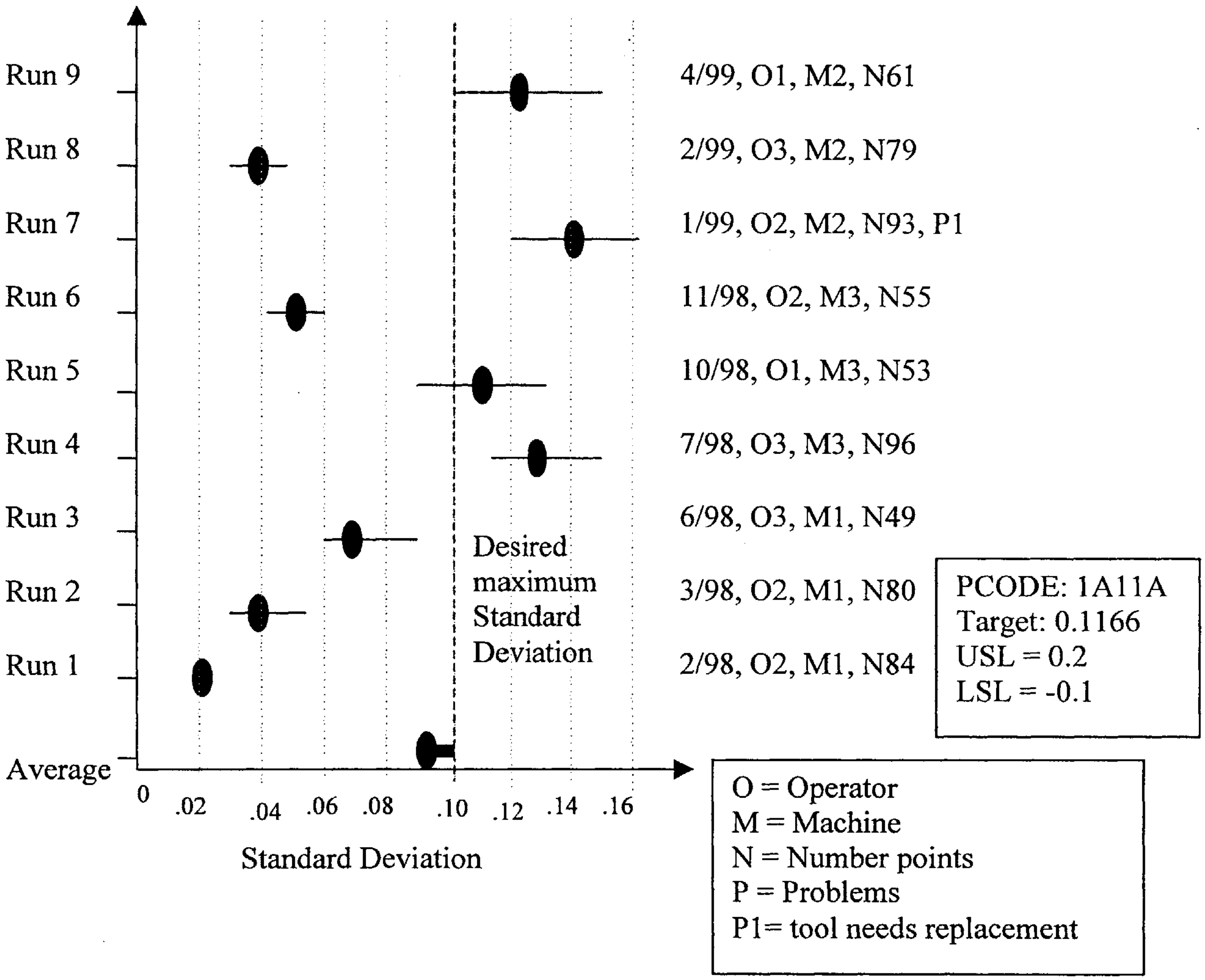


Figure 3.3: Importance of uncertainty in PCD plots

In Figure 3.4, the average point value and the entire confidence interval for the desired index is less than the desired maximum standard deviation; therefore, the designer would assume his/her tolerance is acceptable. However, if the designer maintains his/her desired standard deviation and the product is manufactured in a run similar to runs 4, 5, 7, or 9 of Figure 3.4, he/she will have bad parts produced. Run 7 can be excluded because it had a problem, but runs 4, 5, and 9 cannot be excluded. This shows the importance of allowing the designer to see all of the runs of PCD for the desired index rather than just the average of all the runs.

Overall, the data in Figure 3.4 shows that the designer should increase his/her desired standard deviation value in order to ensure that a sufficient number of quality parts are produced. After eliminating run 7, the maximum confidence interval value for the standard deviation is 0.0852, so the tolerance should be greater than or equal to this value. Alternatively, the designer could specify that Machine 1 always produces this particular part/feature. Machine 1 consistently produces parts with smaller standard deviation values than desired as shown in runs 1, 2, and 3.

Section 5.2.2 provides more details on adding uncertainty to process capability data.



	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8	Run 9	Total
Std Dev	0.02	0.04	0.07	0.13	0.11	0.05	0.14	0.04	0.12	
95% conf int min	0.0174	0.0346	0.0584	0.1139	0.0923	0.0421	0.1224	0.0346	0.1018	
95% conf int max	0.0236	0.0474	0.0875	0.1515	0.1361	0.0616	0.1636	0.0474	0.1461	
conf int length	0.0062	0.0128	0.0291	0.0377	0.0438	0.0195	0.0413	0.0128	0.0443	
average std dev										0.0925
95% conf int min										0.0878
95% conf int max										0.0979
total conf int length										0.0101

Figure 3.4: Importance of plotting all runs for desired PCD index

3.6 *Current state of PCDBs*

The process capability database of the large aerospace company is currently not directly accessible by designers. This is because the database is not simple to use, the data is not in a statistically reliable format, and the database is highly unpopulated. Designers are not provided with direct access to the database because it is feared they may either obtain the wrong data or use the data improperly.

3.6.1 Flow of PCD

Since access to PCD at the large aerospace company is extremely complicated, it was useful to create a drawing of the progression of the PCD between production, quality, and design. This is shown in Figure 3.5. The shapes of the various steps also indicate which group (quality, operations, design, or process capability owners) is responsible for each action.

First, a design has to be made without using process capability data because the data is not originally available. After the part is designed, it is produced. When the part is being manufactured, data on the process capability for that part is collected as part of process control.

PCD should be collected for the various dimensions of the part. The data for each dimension is then entered into the process capability database under an index assigned to it based on its material, process, and feature. Once the data is in the database, the designer can use it when he/she is creating a similar part.

The designer first requests the process capability data for the similar part through either an informal request or a Process Capability Acquisition Request (PCAR). A PCAR is a formal request for the process capability data. When a PCAR is issued, the variation reduction coordinator obtains the necessary data from the process capability database and provides this information to the designer. The variation reduction coordinator (VR coordinator) is someone who is familiar with the process capability database. He/she knows how to determine which data is statistically valid, how to find data for similar parts, and alternative data to use when the

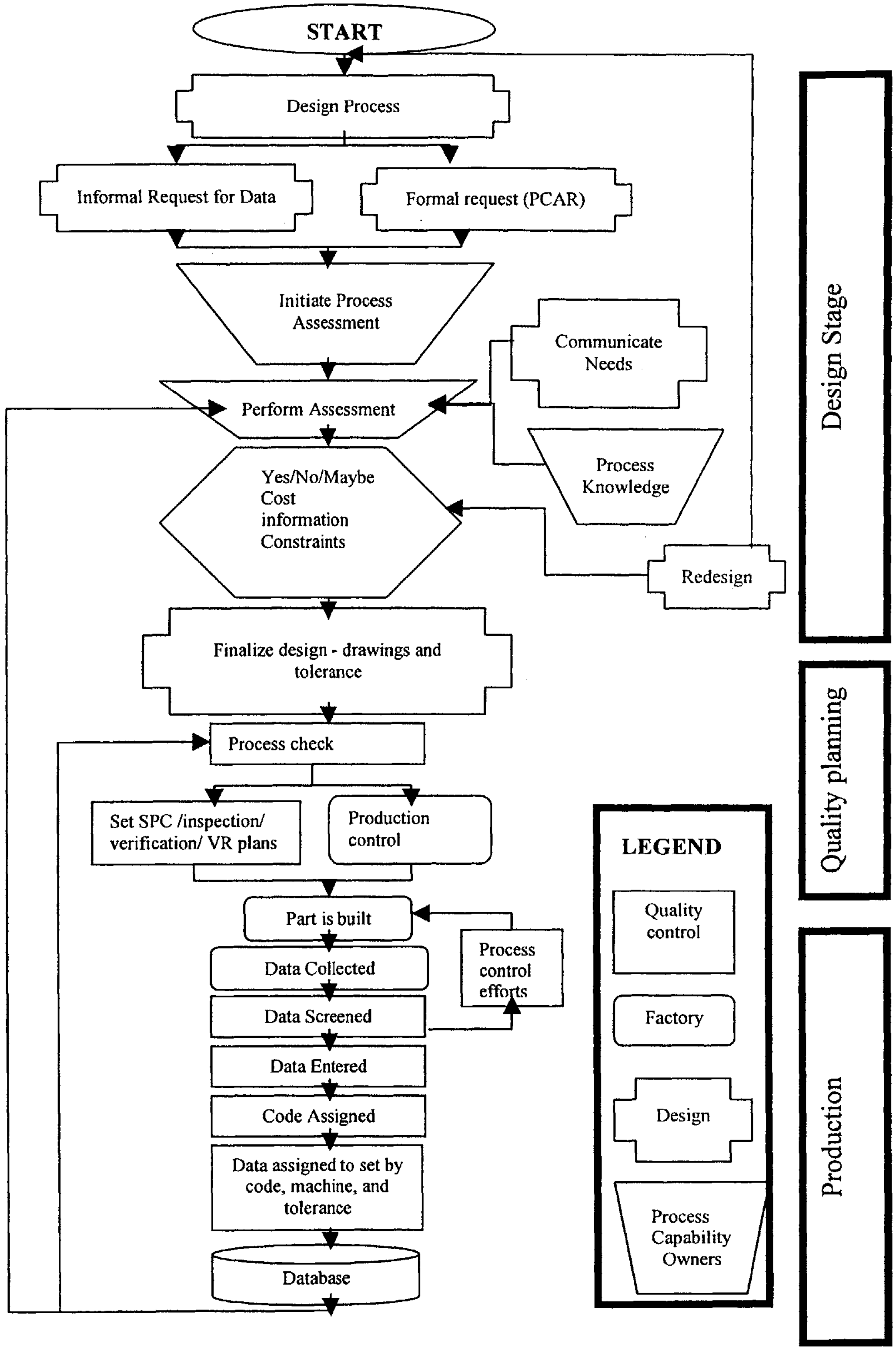


Figure 3.5: Flow diagram of PCD at a large aerospace company

desired data is unpopulated. With the process capability information, the designer modifies his/her drawing. Depending on the severity of the modifications, another PCAR may be used to check the part dimensions. After the design is finalized, it moves into production.

Based on the survey responses discussed in Chapter 2, it seems that production typically does produce the process capability data as shown at the bottom of Figure 3.5. It also appears that quality uses the PCD for process checking and control plans. The problem arises because the design usage of PCD, indicated at the top of Figure 3.5, is not realized to the extent it should be.

3.6.2 PCD access

Rather than accessing PCD directly, designers at the large aerospace company need to request the data through PCARs. There are several problems with using the VR coordinator to obtain the data for the designers. One problem is PCARs take about a week to process. If the designer had direct access to the process capability database, he/she could obtain the data in a matter of minutes. Using the VR coordinators is also a waste of resources. Oftentimes, the designer and VR coordinator have to discuss, at length, which data is needed and for what purpose. If the designer had direct access the time required for this transfer of knowledge could be eliminated.

Finally, it is unclear how VR coordinators determine surrogate and aggregate data to provide to the designer. They do not seem to have any quantitative measures to do this; therefore, they just use their knowledge and experience with the data. The large aerospace company indicated that only 25% of their PCARs are answered and the data used. For the other 75% of the data, either the data is unpopulated, there is not enough information on what data is needed and the wrong data is provided, or the data is not statistically valid and is not used. The enhancements to the large aerospace company's process capability database that are proposed in this thesis should provide some initial steps to allow the designer direct access to the databases.

3.7 Conclusion

This chapter details one of the factors resulting from the communication organizational issue – designer access to PCD. It also deals with the factor of data timeliness, which is part of the organizational issue of a company-wide vision of PCD usage. This chapter explains how the complicated progression of PCD from production to quality to design results in the problems of time lag for designers to obtain PCD and difficult design access to PCD. If these and the other organizational barriers could be eliminated, the various design uses for PCD discussed in Chapter 4 could be achieved.

Based on the industry questionnaire ten issues hampering the use of process capability data (PCD) by the design community were identified. The three organizational issues were discussed in Chapter 2. In addition, further discussions with industry identified seven technical issues.

- PCD is provided as a point value with no measure of uncertainty. Section 5.2.2 explains how confidence intervals enhance the statistical validity of the process capability data by quantifying its uncertainty.
- PCD is presented numerically rather than graphically. PCD is also generally plotted as the average of all the runs and the PCDBs only have the ability to show one run at a time. This inadequate visualization of PCD is discussed in Sections 5.3.1 and 5.4.1. The prototype software developed eliminates this barrier.
- PCDB user interfaces and indexing schemes are difficult to use because there is no consistent PCDB structure. Sections 3.3 and 3.4 explain the hierarchy and indexing scheme for the PCD at the large aerospace company and Section 7.3 presents some possible methods for making it more consistent.
- PCDB user interfaces allow choosing of infeasible indexes. Section 7.2 explains how infeasible index options are presented in the current system and how they can be eliminated.
- PCDBs contain no methods to obtain alternative data when the particular index that the user desires is unpopulated. Section 7.4 details how surrogate data can be chosen for unpopulated indexes.
- PCDBs contain no methods to display aggregate data when designers do not know all the details of the material, feature, and process that they intend to use. Options for aggregate data

to provide to a designer when the specifics of a part are not yet known are detailed in Section 5.4.

- PCDBs contain no methods to eliminate outlier data or to combine runs. Elimination of outliers is discussed in Sections 5.3.2 and 5.4.2. Combining runs and grouping samples are discussed in Sections 5.3.3. and 5.4.3 respectively.

4 Future Process Capability Data Uses and Needs

4.1 Introduction

If the technical issues at the end of Chapter 3 are addressed, designers should be given direct access to PCDBs, so the timeliness and access portions of the organizational issues will be eliminated. Once these issues are addressed, the PCD uses discussed in this chapter will be fully realized. This chapter also details the framework for the steps a designer needs to take to obtain data from a PCDB to determine if a tolerance is acceptable. This framework shows the need for all of the technical issues discussed in Chapter 3 to be addressed.

4.2 PCD uses

Several uses for process capability data were determined by industry contacts. Internal uses include design, day-to-day production, and long-term production. There are also several potential external uses for PCD such as supplier management. In order to obtain all of these PCD uses, there are several PCDB needs for each of supplier management, the customer, quality, manufacturing, and design.

PCD is needed for the design tools of Variation Simulation Analysis (VSA), tolerance analysis and Key Characteristics (KCs). PCD is needed for VSA assessment and tolerance allocation that meets design intent. It is also needed for product verification based upon the capability of KCs, which are used to determine the parameters that need to be measured. It is needed to validate the effectiveness of a KC by determining if the capability of a characteristic yields the desired intent. Finally, it can be used to determine if the capability of a process improves by applying a control plan to it.

PCD can be used in various stages of the design cycle. It should be used initially to create conceptual designs using product architecture. These initial concepts should be deemed manufacturable to the capability of the available machines. Later, PCD can be used as a corrective action to initiate design changes in previous designs that are not manufacturable.

For day-to-day production, PCD can be used for maintenance scheduling including both preventative and long-term. It can also be used for statistical process control, for prevention of process degradation, and process validation. PCD should be used for product verification, variation reduction, and process improvements.

For long-term production, PCD can be used for machine purchasing decisions. It can be utilized to compare rebuild and replace options for existing machines and to determine which machines to acquire to meet future requirements. It can also be used for determination of process capability effect on end-item performance and inspection resource needs.

For production control, PCD is used for the determination of optimal lot sizes based on yield and machine utilization. It is also used to assess parts and for machine allocation. PCD can be used for corrective action through error tracing. For quality control, PCD is used to determine inspection levels and product acceptance. It is also used for operator verification and for monitoring on-going quality. Supplier management uses PCD for supplier selection based upon capability and as input into preferred supplier certification. PCD can be used to determine the appropriate inspection requirements for a given supplier and to validate supplier parts based on historical data.

4.3 PCDB information needs

The PCD needs identified for the design community include standard deviation, mean shift (x-bar from target), relative costs, and gage. Other types of PCD needed by design are performance, risk management, system integration, and flag raising for tolerances that are not process capable. The design community also needs to be able to obtain the process capability data quickly. For this they need a friendly user interface to the database. It is also easier for designers to find PCD indexed by material, feature, and process rather than by part number.

Supplier management or possibly the product buyer would want information on C_p and C_{pk}. C_p is the process capability index. C_{pk} is the process capability. More information on C_p and C_{pk} are contained in Appendix C.

The information the customer would want from a PCDB includes system performance, reliability, and customer-defined capability. Quality would desire control charts and yield specific capability from PCDBs.

Manufacturing currently records several PCD values but they would like to obtain more. For day-to-day production they need moving averages, yield, C_p, C_{pk}, gage, raw data, process, environment, special causes, and tolerances. They need this day-to-day production information on time-based view graphs and control charts. These same data values are needed for long-term production. In order to maintain PCD for long-term production, support, planning, and strategy are needed.

4.4 Ideal state

This section details the framework for the steps a designer would need to take to obtain process capability data. Figure 3.6.2 shows how complicated the process is for a designer to obtain PCD. When a designer specifies a tolerance on a part/feature, he/she uses the PCDB to determine what tolerances were met for past parts with similar characteristics.

The formula for the tolerance obtained for a manufactured part follows:

$$\textit{Manufactured tolerance} = \textit{mean shift} + x(\textit{standard deviation}) \tag{1}$$

Where x is a whole number usually between one and six. The value of x is determined by the desired C_{pk} value. C_{pk} is defined in Section 5.2. The tolerance specified on the design must be greater than the manufactured tolerance in order for the part to pass quality inspection.

$$\textit{Design tolerance} \geq \textit{Manufactured tolerance} \tag{2}$$

The designer would first need to use the prototype software to examine the standard deviation and then he/she would need to use the software to examine the mean shift. If the designer wants a tolerance of 0.01, he/she would first input an upper specification limit of 0.01 and a lower specification limit of (-0.01). Then the designer would determine what amount of this tolerance he/she expects to come from the standard deviation and what amount he/she expects will come from the mean shift.

In most instances, the designer will assume the ideal – that there is no mean shift and will input 0.01 as the desired standard deviation. If the designer queries the PCDB and finds that all the PCD standard deviations are less than 0.01, than the designer progresses to the mean shift. Otherwise, the designer must try to determine if certain control parameters (such as machine or operator) affect the data. If certain control parameter values produce standard deviations consistently less than 0.01, then the designer can specify these parameter values and progress to the mean shift. If control values cannot be specified to produce the desired results, than the designer must return to step one and input a new desired tolerance. Alternatively, he/she can use the maximum standard deviation of the plotted PCD as the tolerance.

Ideally, the mean shift can be controlled to zero and the desired tolerance of 0.01 should be met if all the standard deviation PCD is less than 0.01. However, it is also useful to look at the mean shift to make sure it is small and its summation with the standard deviation produces a value less than the tolerance. The designer progresses through the identical process for the mean shift as was used for the standard deviation. The desired mean shift input by the designer would be the difference between the maximum standard deviation obtained from the PCD and the desired tolerance.

As shown by Figure 4.1, the process of using PCD to determine if a tolerance is acceptable is complicated. First the designer must input the upper and lower specification limits, and the dimension. If the designer knows the index desired, this is also input.

If the desired index is feasible, then the designer inputs the desired tolerance and either single or multiple runs are plotted. Since the index is composed of material, feature, and process, an index is feasible if the combination of that set of material, feature, and process value both can be combined and has been combined in the past. For example, drilling a hole in aluminum is feasible but injection molding a hole in aluminum is not feasible.

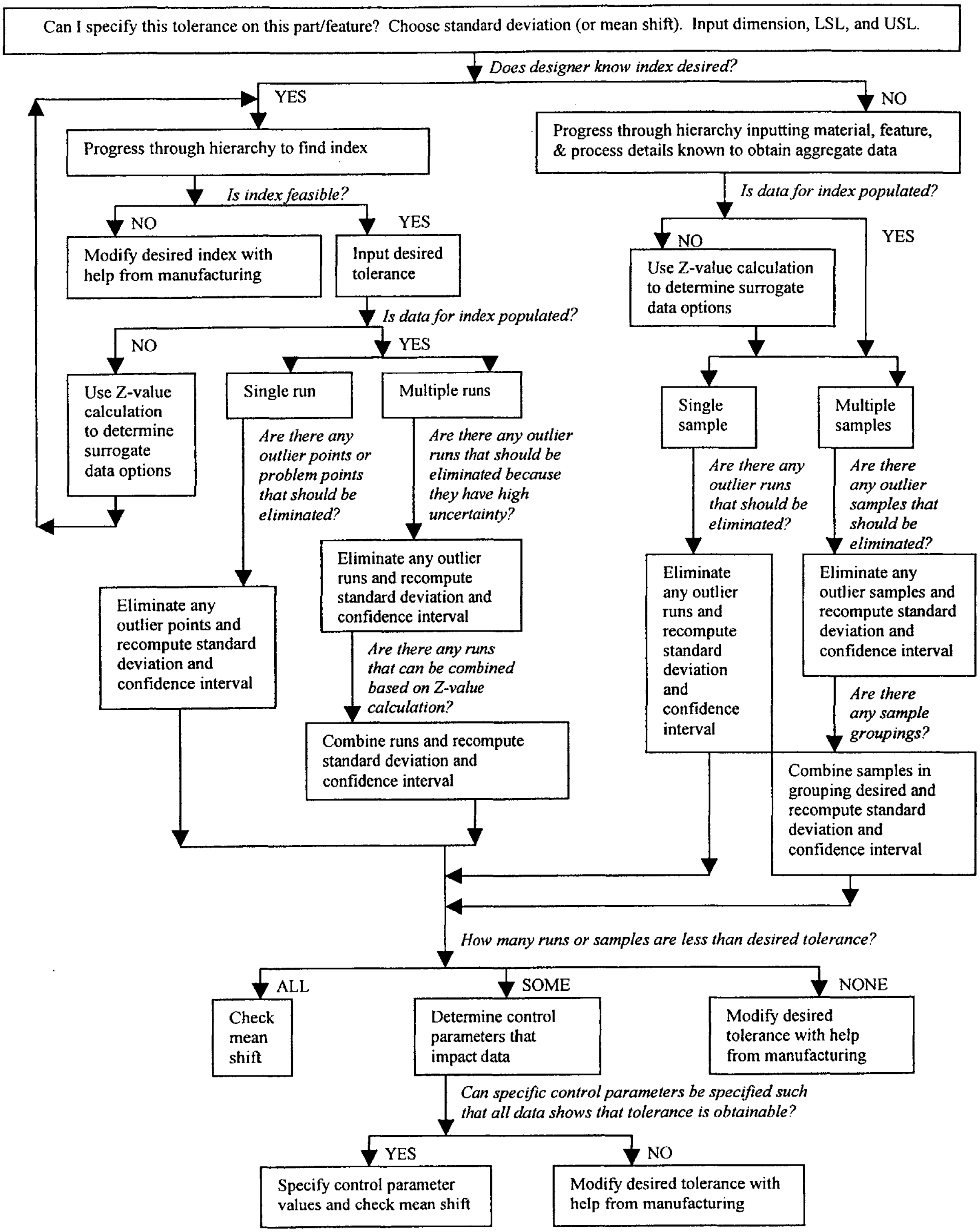


Figure 4.1: Flow diagram of framework for designer use of PCDB

Next, outlier points or runs must be eliminated and the standard deviation and confidence interval values must be recomputed. Details on how to determine if runs are outliers are provided in Section 5.3.2. If there are multiple runs, the Z value calculation can be used to determine if any of the runs can be combined. If some runs are combined, the standard deviation and confidence interval values must be recomputed. The Z value calculation is a quantitative method to compare two runs to see if they are similar. Details on the Z value calculation are provided in Section 5.2.3.

To determine how many runs are less than the desired tolerance, the designer looks at the plot to see how many of the runs satisfy the desired tolerance. The desired tolerance is plotted as a vertical line. If all the data is to the left of this line, the desired tolerance will be acceptable based on the standard deviation; therefore, the designer should next check the mean shift. If none of the data is to the left of the desired tolerance, then the designer must work with manufacturing to modify the desired tolerance.

If only some of the data is to the left of the desired tolerance line, the designer should examine the labels for each run of data to determine if all values to the right of the line are of a certain control parameter value. A control parameter might be a machine or an operator. If just one control parameter value is to the right of the desired tolerance line, then this value can be eliminated and the desired tolerance is acceptable. In this case, the designer must specify that the designated control parameter be avoided for this tolerance. If the control parameter values vary such that all possible values have at least one run of data to the right of the desired tolerance line, then the designer must work with manufacturing to modify the desired tolerance.

If the designer had chosen a index that was not populated, he/she would need to use the Z value calculation to determine aggregate data. Once this data was determined, he/she would need to return to progressing through the hierarchy to find the desired aggregate data. Section 5.4 further details the process of obtaining aggregate data. If the index selected was not feasible, the designer must determine a feasible index for which to obtain data. Section 7.2 provides information on infeasible indexes.

If the designer had not originally known the index desired, he/she would need to progress through the hierarchy inputting the details he/she knew. These details would result in either a single or multiple samples. Outlier runs or samples would have to be eliminated and the standard deviation and confidence interval values recomputed. If there were multiple samples, groupings could be determined and the standard deviation and confidence interval values would be computed for the particular grouping(s) chosen by the designer. Finally, the designer would be able to determine how many samples were less than the desired standard deviation. Chapter 5 further develops the details of Figure 4.1 such as aggregate data, determining outlier data, grouping samples, and combining runs.

There are several features of the framework of Figure 4.1 that are not addressed in current industry PCDBs, but which are discussed in this thesis.

- First, the ability to see multiple runs of data at one time (Section 3.5).
- Second, the ablity to determine and exclude outlier points for a run and then recompute the average standard deviation for the run (Section 5.3.2).
- Third, the ability to determine and exclude outlier runs for a sample and then recompute the average standard deviation for the sample (Section 5.3.2).
- Fourth, the ability to determine and exclude outlier samples for a set and then recompute the average standard deviation for the set for obtaining aggregate data (Section 5.4.2).
- Fifth, the ability to input a desired tolerance and graphically compare it to the standard deviation or mean shift of all the runs for a particular indx, machine, target value, and range of specification limits (Section 5.5.4).
- Sixth, the ability to see multiple samples of data at one time to determine groupings for aggregate data (Section 5.4.3).
- Seventh, the ability to combine samples in a grouping for aggregate data (Section 5.4.3).
- Eighth, the ability to quantitatively determine if runs can be combined (Sections 5.2.3 and 5.3.3).
- Ninth, the ability to quantitatively determine if one run is an acceptable surrogate for another run (Section 7.4.2).

- Tenth, the ability to know the range of values for an index with a 95% confidence interval (Section 5.2.2).

4.5 Conclusion

The framework for the steps that a designer must take to obtain PCD is provided in this chapter. This framework shows the need for the seven technical barriers for design usage of PCD to be eliminated. Ten features that are missing from but needed in current PCDBs were identified. Means of adding these features are detailed in the remainder of this thesis.

5 Visualization of Process Capability Data

5.1 Introduction

Chapter 4 discussed the needs and potential uses for process capability data. A prototype software system was developed to overcome the two main issues hampering the use of process capability data by design. These issues are the statistical validity and the visualization of process capability data.

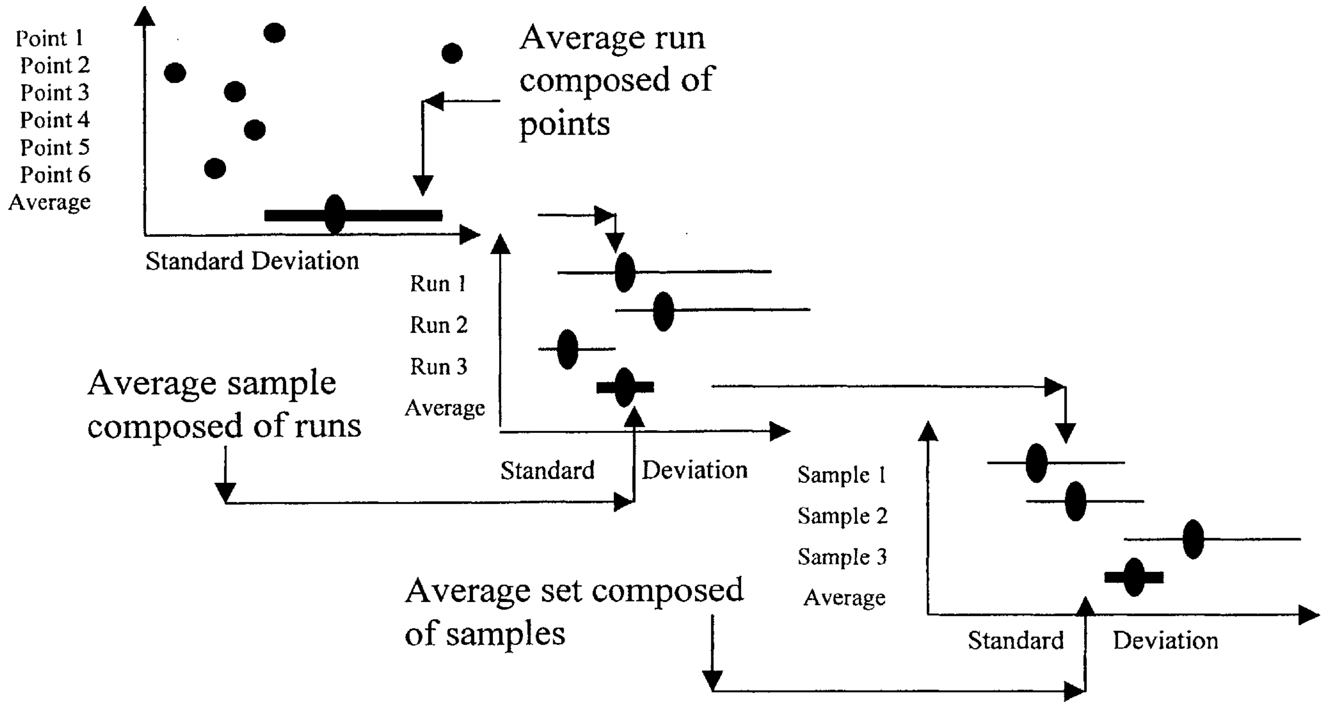
The issue of statistical validity is addressed through three means. The first is adding confidence intervals to the PCD rather than just providing a point value (Section 3.5). By providing a range of values with some confidence, the uncertainty of the data is quantified. The second is providing the values for each run in a sample rather than just providing the average value for all the runs. The third is using a visual analysis to determine outlier data, which should be excluded (Sections 5.3.2 and 5.4.2). After particular data has been excluded, it can be quantitatively determined which runs can be combined (Sections 5.2.3 and 5.3.3). It can also visually be determined which samples can be grouped (Section 5.4.3). This eliminates the possibility of incompatible data being combined.

Uncertainty is a function of the number of data points collected and the variation between these data points. Uncertainty also depends on the transient behavior of the data and on whether the data is the actual data desired or surrogate data. Uncertainty is higher when there are fewer data points, when surrogate data is used for unpopulated indexes, or when there are frequent process changes that affect the data. By including the uncertainty with the data, the designer will better understand its accuracy and thereby have a higher level of confidence.

There are two types of data that need to be plotted. One type is a set of runs for a particular index (material, process, and feature) – Section 5.3.1. The other type is a set of samples for when the designer only knows two of the three parameters of feature, material, and process. This set of samples is termed "aggregate" data and is discussed in Section 5.4.1.

There will be a level of uncertainty between each run, which has several of data points; between each sample, which has many of runs; and between each set, which has a plethora of samples. There may also need to be normalized uncertainty between surrogate data used for unpopulated indexes. The uncertainty progression from point to run to sample to set is shown in Figure 5.1.

Each point in a run will have the same index, target dimension, specification limits, machine, date and operator. This set of points will be combined to determine an average standard deviation with a particular confidence interval for the run. The run will then be combined with several other runs to form a sample. The sample will be composed of runs that all have the same index, target dimension, and specification; however, the runs will have varying machine, date, and operator. The set of runs will be combined to determine an average standard deviation with a particular confidence interval for the sample. The sample will then be combined with several other samples to form a set. The set of samples will be composed of samples that have the same target dimension and specification limit. The samples will also all be of the same feature and material, material and process, or feature and process.



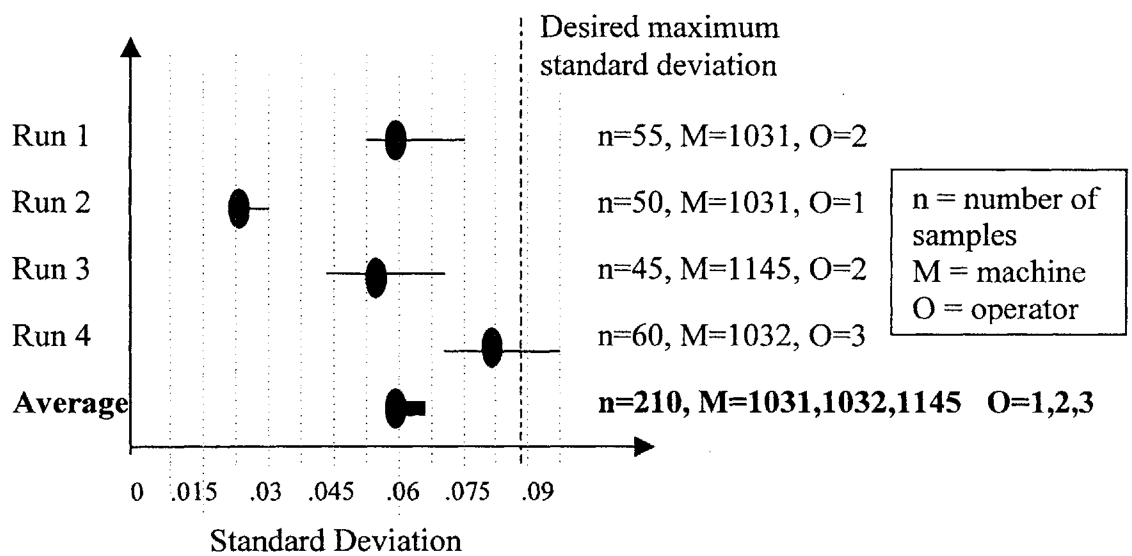
RUN 1	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6	Total no.
Std dev	0.03	0.08	0.01	0.02	0.025	0.015	6
Average std dev							0.03797
95% conf int min							0.0237
95% conf int max							0.09314
conf int length							0.06943

SAMPLE 1	Run 1	Run 2	Run 3	Total
Std dev	0.038	0.05	0.02	
No. points	6	8	12	26
95% conf int min	0.024	0.033	0.014	
95% conf int max	0.093	0.102	0.034	
conf int length	0.069	0.069	0.020	
average std dev				0.034
95% conf int min				0.027
95% conf int max				0.047
conf int length				0.020

SET 1	Sample 1	Sample 2	Sample 3	Total
Std dev	0.034	0.04	0.06	
No. points	26	40	36	102
95% conf int min	0.02666	0.03277	0.04867	
95% conf int max	0.04693	0.05136	0.07827	
conf int length	0.02027	0.0186	0.0296	
average std dev				0.046
95% conf int min				0.041
95% conf int max				0.054
conf int length				0.013

Figure 5.1: Depiction of uncertainty chain from point to run to sample to set

The issue of visualization is addressed in the prototype software by graphically displaying the data. The standard deviation or the mean shift of each run is displayed as circle and the confidence interval is displayed as a line passing through this circle. Each run is plotted on an individual line, which is labeled with important information about the data, so the designer can recognize patterns. Figure 5.2 shows an example of what this graphical representation might look like for fictitious data.



SAMPLE	Run 1	Run 2	Run 3	Run 4	Total
Std Dev	0.06	0.025	0.055	0.08	
No. points	55	50	45	60	210
95% conf int min	0.05051	0.02088	0.04553	0.06781	
95% conf int max	0.07392	0.03116	0.06949	0.09758	
Conf int length	0.02341	0.01028	0.02396	0.02977	
average std dev					0.0593
95% conf int min					0.0541
95% conf int max					0.0656
conf int length					0.0115

Figure 5.2: Basic visualization of data in prototype software

The maximum desired standard deviation in Figure 5.2 would be input by the designer and would be equivalent to the tolerance for the part/feature. In order for this tolerance to be acceptable, all standard deviation confidence intervals for the index plotted must be less than the maximum desired standard deviation value.

The calculations used in the prototype software are detailed in Section 5.2.1. Section 5.5.1 explains the structure of the portion of the large aerospace company PCDB used in the software. Section 5.5.2 details the methodology of the software, which is composed of a user interface (Section 5.5.3) and a data output form (Section 5.5.4). The software generates plots similar to the one shown in Figure 5.2 for both mean shift and standard deviation. Examples of the features of the software are provided for mean shift and for standard deviation in Chapter 6.

5.2 Theory

The prototype software performs several mathematical analyses in order to enhance its statistical validity. These computations of mean shift, specification limits, and C_{pk} are discussed in Section 5.2.1. Confidence intervals are also included in the prototype software to minimize data uncertainty (Section 5.2.2). Finally, Section 5.2.3 details a quantitative analysis which can be used to determine if two runs can be combined.

5.2.1 Software computations

The sample database provided by the large aerospace company contained the following data: PCODE, lower tolerance, upper tolerance, target, machine, and measurement. With this data as input, several formulas are used in the prototype software. These are used to generate columns in an ACCESS database. More information about the use of this database in conjunction with the prototype software is provided in Sections 5.5.1 and 5.5.2. The formulas follow:

$$\textit{Lower specification limit} = LSL = \textit{lower tolerance} - \textit{target} \tag{1}$$

$$\textit{Upper specification limit} = USL = \textit{upper tolerance} - \textit{target} \tag{2}$$

Where the *target* is the value for a dimension specified on a drawing. The *lower tolerance* is the minimum value for a dimension specified on a drawing. The *upper tolerance* is the maximum value for a dimension specified on a drawing. When the part is manufactured it is desired that the measurement of the dimension to be equal the *target*; however, in order to be accepted, the measurement must be greater than the *lower tolerance* and less than the *upper tolerance*. The tolerances and targets are depicted in Figure 5.3.

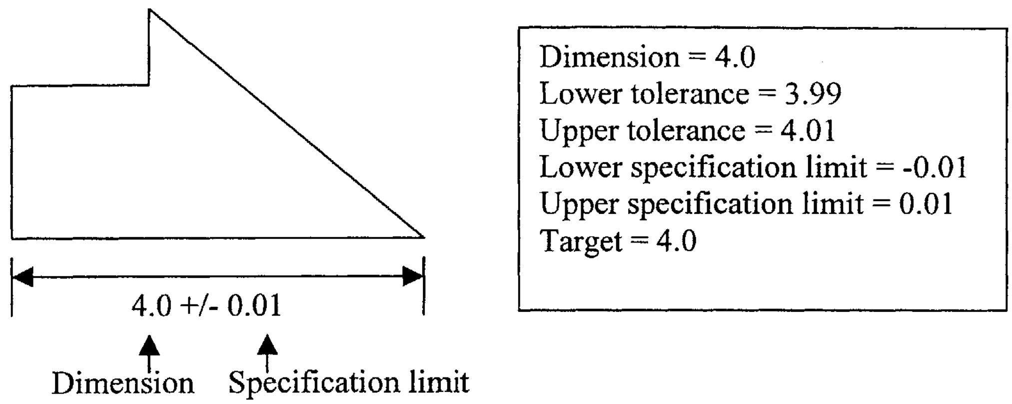


Figure 5.3: Depiction of tolerances, target, and specification limits

$$Number\ of\ points = n = count(*) \tag{3}$$

In ACCESS, equation 3 is used to count the number of points for a particular set of sorting parameters. The data is "grouped by", or sorted by, PCODE, target, LSL, USL, and machine id. Each group is considered a run. Since operator and date labels were not provided by the large aerospace company, the runs cannot be separated by these features. *n* is the number of points in each run.

$$Mean = Average = \bar{x} = \frac{X1 + X2 + ... + Xn}{n} \tag{4}$$

Where *X1, X2, ... Xn* are the values of each point in the run and *n* is the total number of points in the run.

$$Variance = \sum_{i=1}^{n} \frac{(X_i - \bar{x})^2}{n-1} = \sigma^2 \tag{5}$$

$$Standard\ Deviation = \sigma = \sqrt{\sum_{i=1}^{n} \frac{(X_i - \bar{x})^2}{n-1}} \tag{6}$$

Where X_i is the value of each point in the run, n is the total number of points in the run, and $\bar{x}$ is the mean for the entire run.

$$Average\ Standard\ Deviation = \sigma_T = \sqrt{\frac{\sum_i (\sigma_i)^2 (n_i - 1)}{\sum_i n_i - 1}} \tag{7}$$

$$Mean\ shift = M = average\ value - target \tag{8}$$

$$C_{pk} = Minimum\left(\frac{M - LSL, USL - M}{3\sigma}\right) \tag{9}$$

Where σ_i is the standard deviation of each run and n_i is the number of points in each run.

5.2.2 Uncertainty

Uncertainties minimize data validity. There are a variety of uncertainties in process capability databases including surrogate data, multiple data sets, aggregate data, and small data sets. There are various other uncertainty causes according to Thornton (working paper):"…processes degrade over time, suppliers change, and unexpected problems occur ." As a result, it is impossible to state "the process capability of X has a standard deviation of Y." To be statistically correct, the information derived from the database should be stated as: *there is a 95% confidence that the standard deviation ranges between 0.030 and 0.040 with an expected value of 0.035.*

Uncertainty is included in the prototype software by adding confidence interval ranges. A confidence interval is "an interval of plausible values for the parameter being estimated" (Devore, 1987). The formulas for the upper and lower confidence interval for the mean shift and standard deviation follow:

$$\textit{Lower confidence interval for mean shift} = M - \frac{1.96\sigma}{\sqrt{n}} \tag{10}$$

$$\textit{Upper confidence interval for mean shift} = M + \frac{1.96\sigma}{\sqrt{n}} \tag{11}$$

$$\textit{Lower confidence interval for standard deviation} = \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2,\,n-1}}} \tag{12}$$

$$\textit{Upper confidence interval for standard deviation} = \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\,n-1}}} \tag{13}$$

Where s^2 is the sample variance, n is the number of samples, M is the mean shift of the sample, and σ is the standard deviation of the sample. The confidence interval for the mean shift assumes a normal distribution. This assumption is valid for runs with at least 30 points; however, in this thesis it is used for runs with any number of points. Appendix D shows the portion of the standard normal curve area table for confidence intervals between 90% and 99%. This table is used for the mean shift confidence interval calculations.

The confidence intervals for the standard deviation are based on the chi-squared distribution. $\chi^2_{\alpha,\nu}$ is the chi-squared distribution, which is not symmetric, but which becomes more symmetric as ν increases. The number of degrees of freedom is represented by ν and is equivalent to *(n-1)*. α is the area under the chi-squared curve that lies to the right of $\chi^2_{\alpha,\nu}$. For a 95% confidence interval, $\alpha/2$ is 0.025 and $(1-\alpha/2)$ is 0.0975. Appendix E shows the table of critical values for the chi-squared distribution. This table shows values for $\chi^2_{\alpha,\nu}$ when $\nu < 40$. For $\nu > 40$, the following formula is used to determine the value of $\chi^2_{\alpha,\nu}$ (Devore, 1987):

$$\chi^2_{\alpha,\nu} = \nu\left(1 - \frac{2}{9\nu} + z_\alpha\sqrt{\frac{2}{9\nu}}\right)^3 \tag{14}$$

5.2.3 Quantitative analysis for combining runs

One quantitative method has been developed to determine if two runs can be combined. This Z value calculation can be used to determine if two runs are similar with a 95% confidence interval. The calculation of this Z value follows:

$$Z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \tag{15}$$

Where $\bar{x}$ is the average value for run 1, $\bar{y}$ is the average value for run 2, m is the number of points in run 1, n is the number of points in run 2, σ_1 is the standard deviation of run 1, and σ_2 is the standard deviation of run 2.

For a 95% confidence interval, if $Z > 1.96$ or $Z < (-1.96)$ then run 2 is similar to, and can be combined with, run 1. If Z is not between -1.96 and 1.96 then the data for run 1 should not be combined with run 2.

The way that the 1.96 value is derived for a 95% confidence interval is as follows:
"Because the area under the standard normal curve between -1.96 and 1.96 is 0.95, the following probability statement is valid:" (Devore, 1987)

$$P(-1.96 < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < 1.96) \tag{16}$$

Or after some manipulation:

$$P(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}) \tag{17}$$

Where $\bar{x}$ is the average of run, σ is the standard deviation of run, n is the number of points in run, μ is the predicted average of future runs, and P is probability.

For a 95% confidence interval $\alpha = 1\text{-}0.95 = 0.05$ so $Z(\alpha/2) = Z(0.025)$. This Z value can be looked up in the table of standard normal curve areas, which is in Appendix D. One must look for the value of 0.025 in the table. The value is found at (–1.96); therefore, 1.96 is the value of $Z(\alpha/2)$ for a 95% confidence interval. If a 99% confidence interval was desired, then $\alpha = 1\text{-}0.99 = 0.01$ so $Z(\alpha/2) = Z(0.005)$. The table shows that the value of 0.0051 is at (–2.57) and the value of 0.0049 is at (–2.58); therefore, 2.575 is the value of $Z(\alpha/2)$ for a 99% confidence interval based on interpolation. Finally, $Z(\alpha/2)$ for a 90% confidence interval is 1.645. This Z value calculation is only valid for samples with a normal distribution. Generally a normal distribution can be assumed for samples composed of at least 30 runs. Two examples of using the Z value calculation to determine which runs can be combined are detailed in Section 5.3.3.

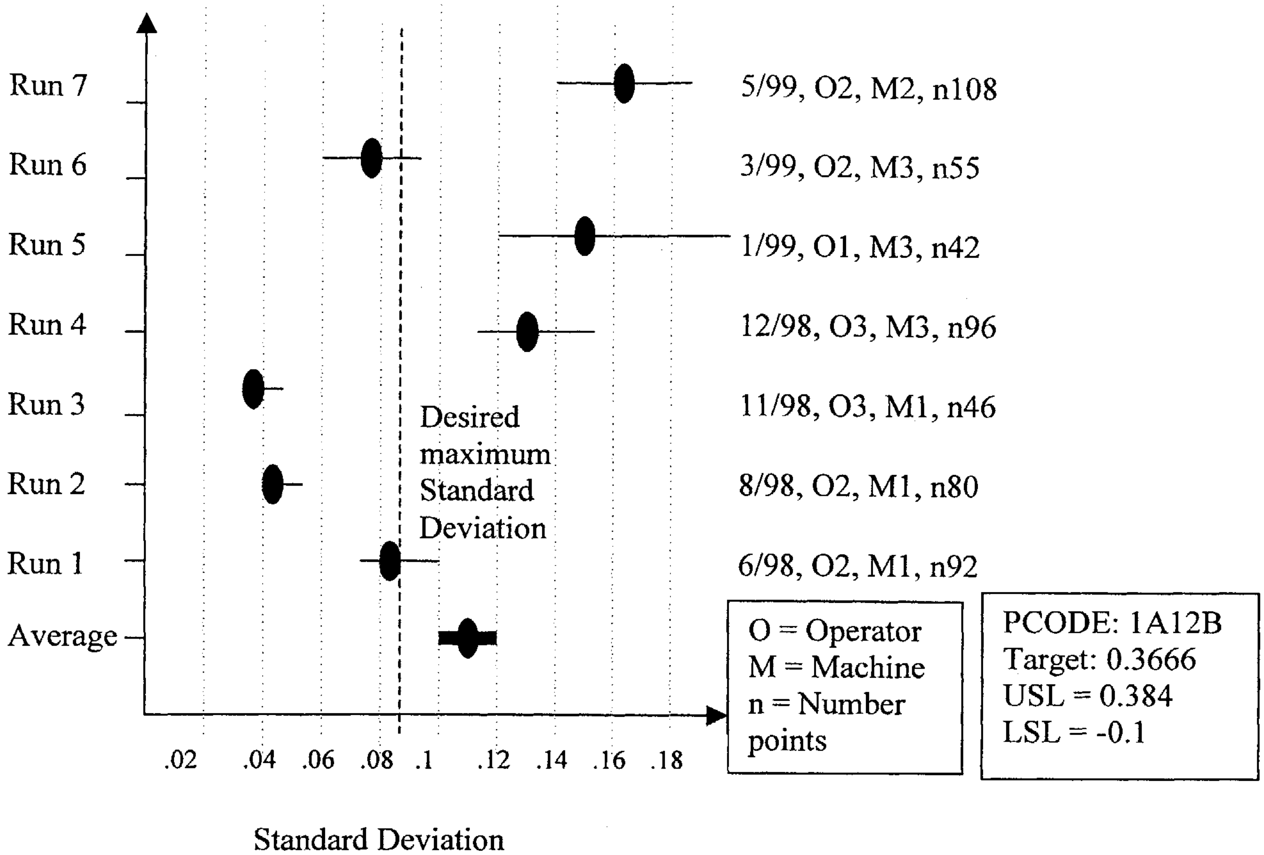
5.3 Numerous runs

A run is made up of a set of points with the same index (PCODE), target value, specification limits, operator, machine, and date. These runs can be compiled to form a sample for a particular index, and set of specification limits, machines, and target values. Alternatively, a set of runs can be used to compile a sample for a particular target value, set of specification limits, and two of the three parameters of material, feature, and process. The machine will vary with the parameters chosen. Section 5.3.1 first details how runs are plotted with a confidence interval. Then, Section 5.3.2 explains how outliers in a run and in a sample can be detected. Finally, Section 5.3.3 provides two examples of using the Z value calculation detailed in Section 5.2.3 to determine if runs can be combined.

5.3.1 Plotting data runs

Figure 5.4 shows an example of plotting the various runs for a fictitious sample of data. In Figure 5.4, the circles represent the average value for the standard deviation and the lines for each run indicate the lower and upper confidence interval for the standard deviation. Since the data shown in Figure 5.4 are for one index, the standard deviation values should all be similar. This figure also shows what it might look like to have labels with each run on the plot. These labels of date, operator, machine, number of samples, and problems aid in the understanding of the data and allow for comparisons.

The "desired standard deviation" in Figure 5.4 would be input by the designer. This vertical line allows him/her to easily see which runs produced standard deviations greater than the value desired. This is helpful in determining if a drawing tolerance needs to be increased. The bottom line in the plot of Figure 5.4 is the average standard deviation range.



SAMPLE	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Total
Std Dev	0.085	0.045	0.035	0.13	0.15	0.075	0.16	
No. points	92	80	46	96	42	55	108	609
95% conf int min	0.074	0.039	0.029	0.114	0.123	0.063	0.141	
95% conf int max	0.099	0.053	0.044	0.152	0.191	0.092	0.185	
conf int length	0.025	0.014	0.015	0.038	0.068	0.029	0.044	
Average std dev								0.112
95% conf int min								0.105
95% conf int max								0.119
Total conf int length								0.014

Figure 5.4: Standard deviation range visualization for multiple runs

5.3.2 Excluding data runs

There are two causes for run or point data to be excluded. The first cause is if the data is an outlier caused by a problem that has been eliminated. The second cause is if the run data both deviates greatly from the average value for the sample and has a high uncertainty.

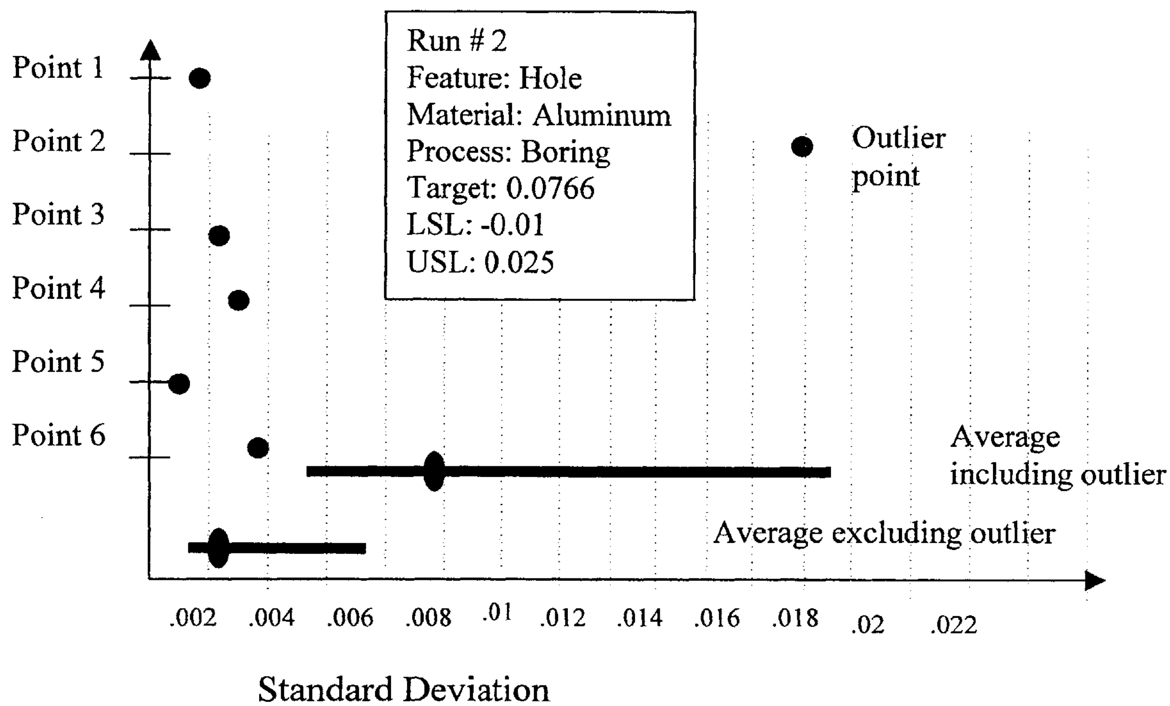
For a stable process, outliers result from variation caused by the operator, gaging system, and the environment, which all may vary. Outliers are data that are substantially different from the rest of the data. Outliers can occur in points, runs, or samples. The average value of the data excluding the outlier is much less or much greater than the value for the outlier.

For this thesis, a run/sample is considered to be an outlier if its upper confidence interval is three times or more greater or its lower confidence interval is 1/10 times or more less than the average point value for the sample/set when the outlier excluded. The outlier run/sample needs to be excluded from the average point value for the comparison because outliers can skew the average greatly such that the average may be closer to the outlier value. For a run, a point is considered an outlier if its value is 300% or more times greater or 10% or more times less than the average point value for the run when the point is excluded.

The following is an example of determining if a run is an outlier. If the average point value for all the runs (excluding the outlier) in a sample is 0.025, then any run with an upper confidence interval that exceeds 0.075 or a lower confidence interval that is less than 0.0025 would be considered an outlier run. Different quantitative methods could be developed to determine outliers; however, for consistency, the aforementioned assumption is used throughout this chapter.

When a point outlier occurs in a run, it should not be excluded unless there is an exact cause for that point deviation. For example, if an outlier occurs because a tool broke, then this outlier should be excluded from the average value. Alternatively, if an outlier occurs because a process is going outside the control limits, it should not be eliminated. This is because it is possible for the process to go beyond the control limits again; therefore, the designer needs to account for this possibility. If action is taken to ensure the process does not go beyond the control limits, then the point outlier value can be excluded.

Figure 5.5, which uses fictitious data, shows what a point outlier might look like for a run. It shows that the average standard deviation for the run is dramatically different depending on whether the outlier is included or excluded. It also shows that the uncertainty is greater when the outlier is included. Figure 5.5 shows that the outlier point value of 0.018 is more than 300% greater than the average point value for the run excluding the outlier, 0.0049. The outlier point value of 0.018 is not more than 300% greater than the average point value for the run including the outlier, 0.00765 but is more than 300% greater than the average point value for the run excluding the outlier, 0.00232. This shows why the aforementioned outlier test uses the average point value excluding the outlier rather than just the average point value.



	Point 1	Point 2	Point 3	Point 4	Point 5	Point 6	Total no.
Std dev	0.0018	0.018	0.0023	0.0027	0.00097	0.0032	6
Average with outlier							0.007649
95% conf int min with outlier							0.004774
95% conf int max with outlier							0.018761
conf int length with outlier							0.013987
Average without outlier							0.002324
95% conf int min w/o outlier							0.001392
95% conf int max w/o outlier							0.006681
conf int length w/o outlier							0.005288

Figure 5.5: Depiction of outlier point for a run

Figure 5.5 also shows that if an outlier point is much larger than the rest of the points for the run, then the length of the confidence interval is usually much shorter when the outlier is excluded than when it is not.

After it has been determined which outlier points within a run should be excluded, the run can be included in a sample of various runs. Again there can be outliers in this comparison of runs. When an outlier occurs in one of the runs for a sample, the outlier should be separated from the rest of the data and possibly excluded.

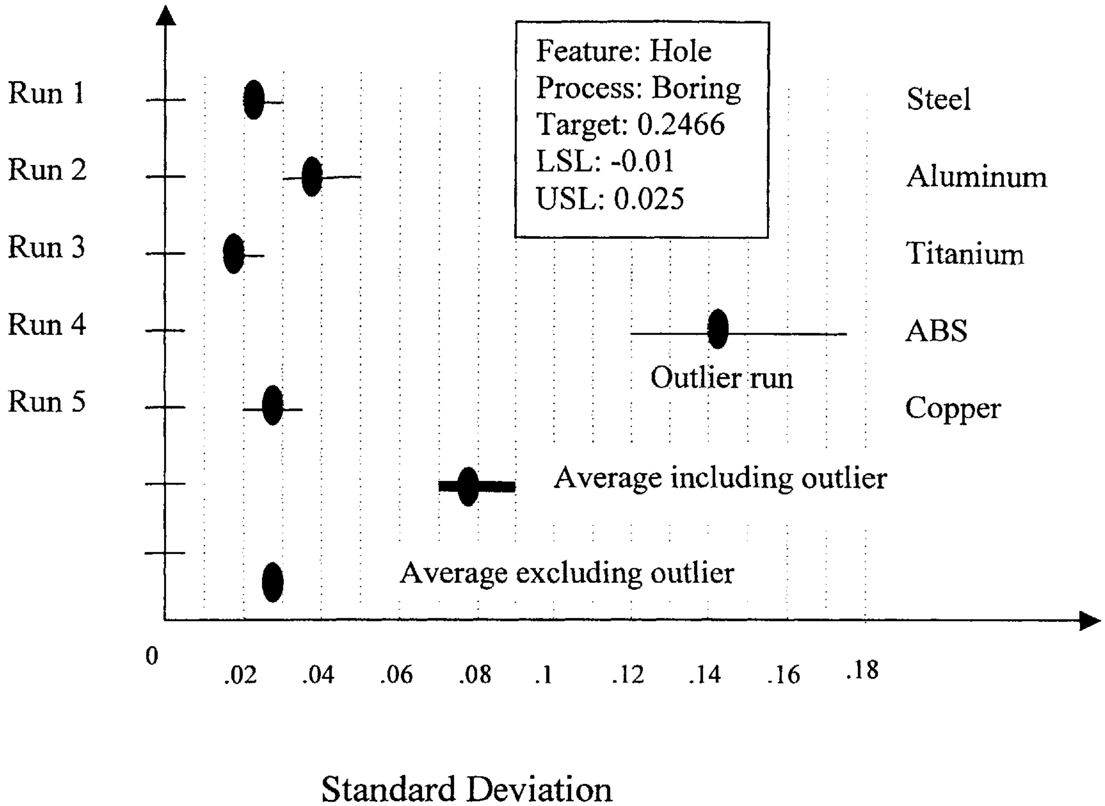
Outlier runs should be excluded from the average value for the sample only if there is a reason for their occurrence. For example, if an outlier occurs because a run was made on an old machine that has since been eliminated, then this outlier should be excluded from the average value. Alternatively, if an outlier occurs because a tool is worn, it should not be eliminated. This is because it is likely for the tool to always be worn for a few parts before it is replaced; therefore, the designer needs to account for this possibility.

If an average value for this sample of runs is desired, it is necessary to determine which data is the most reliable. If data for a particular run deviates greatly from the average and has a high uncertainty, the data isn't sufficiently reliable. However, if the same high deviation data has a low uncertainty, the data is reliable. If data for a particular run has a high uncertainty but does not deviate significantly from the average, then the data is reliable. This is summarized in Table 5.1.

	High Uncertainty	**Low Uncertainty**
Large deviation of run from average	Data unreliable = exclude data	Data reliable
Small deviation of run from average	Data reliable	Data very reliable

Table 5.1: Determination of data reliability

Figure 5.6 shows what a run outlier might look like for a sample. Again this figure includes fictitious data.



SAMPLE	Run 1	Run 2	Run 3	Run 4	Run 5	Total
Std Dev	0.022	0.037	0.018	0.143	0.027	
No. points	55	50	45	60	70	225
95% conf int min	0.01852	0.03091	0.0149	0.12121	0.0231	
95% conf int max	0.0271	0.04611	0.02274	0.17443	0.0324	
Conf int length	0.00858	0.01521	0.00784	0.05322	0.0093	
Average standard deviation w. outlier	0.00086	0.00152	0.00078	0.00532	0.0009	0.078
95% conf int min						0.0714
95% conf int max						0.086
conf int length						0.0146
Average standard deviation w/o outlier						0.0266
95% conf int min						0.0243
95% conf int max						0.0293
conf int length						0.005

Figure 5.6: Depiction of outlier run for a sample

Figure 5.6 shows that if an outlier run is much larger than the rest of the runs for the sample, then the length of the confidence interval is usually much shorter when the outlier is excluded than when it is not.

The fictitious data in Figure 5.7 shows a run comparison that can be used to determine which data is not reliable and should be excluded.

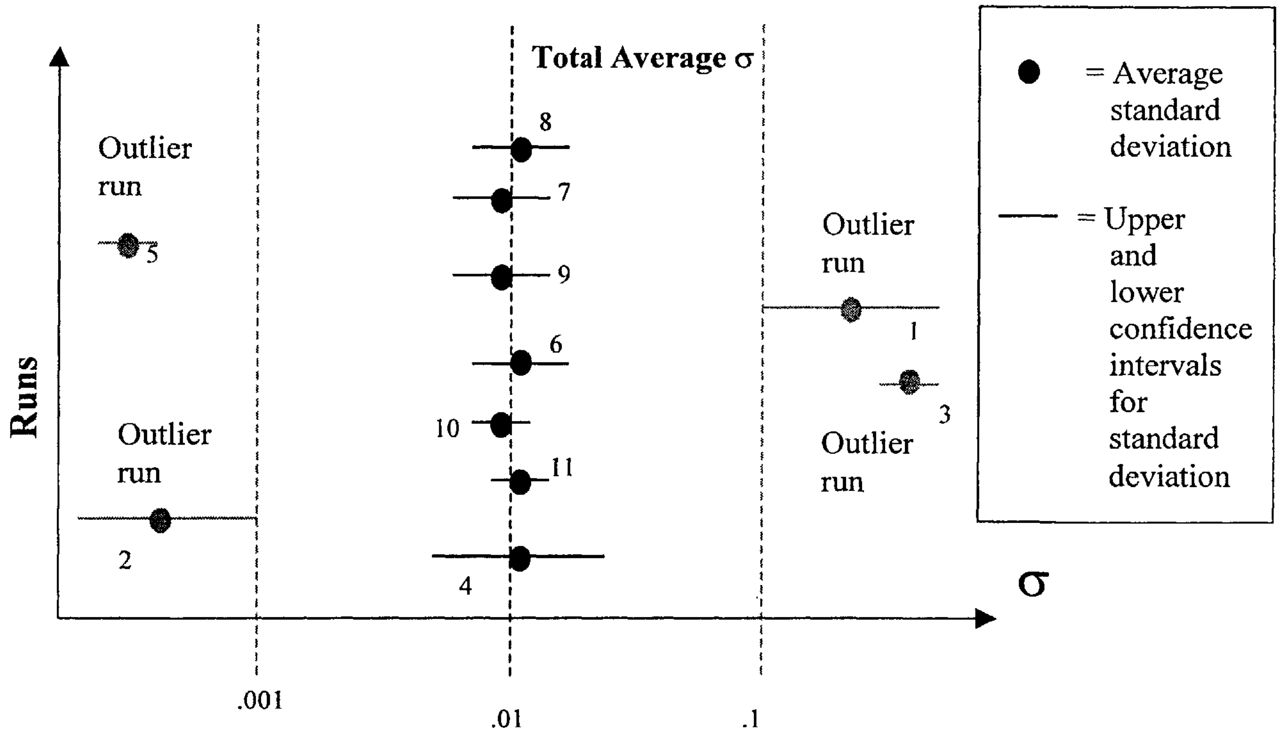


Figure 5.7: Plot to determine which outlier runs should be eliminated

In Figure 5.7, run 1 consistently has a standard deviation significantly greater than the average; however it also has a lot of uncertainty for this standard deviation, as shown by the length of the line representing its upper and lower confidence intervals. Run 1 should be excluded from the average. Run 2 has a standard deviation significantly less than the average and has a lot of uncertainty; therefore, it should also be excluded.

In Figure 5.7, run 3 has a standard deviation significantly greater than the average with a small amount of uncertainty. This means run 3 should not be excluded because of the high certainty of the data. Run 5 has a standard deviation significantly less than the average with a small amount of uncertainty; therefore, its data should not be excluded. Finally, run 4 has an average standard deviation similar to the total average standard deviation; however it has a lot of uncertainty. Because the significant uncertainty still results in an average similar to the total average, run 4 should not be eliminated.

Overall, data should be eliminated only if it has a high uncertainty and its standard deviation is not similar to the average standard deviation. Therefore, all runs except runs 1 and 2 should be used to calculate the total average for all the runs. A plot like that in Figure 5.7 makes it easy to visually determine outlier runs. The average value obtained by excluding runs 1 and 2 has a much higher validity than the original average shown in Figure 5.7.

It is unclear how the run data should be averaged for the sample. For example, in Figure 5.7, should the data for all runs except 1 and 2 simply be added and divided by 9 (the total number of runs excluding 1 and 2) or should the sample size and uncertainty for each run also be included in the calculation. It seems that runs with less uncertainty and/or runs with higher sample sizes should be given more dominance in the calculation. Nonetheless, in this thesis the first method is used.

5.3.3 Combining data runs

The process capability data can be simplified by combining runs that have the same index (this encompasses material, feature, and process), target value, and specification limits. An average value and a standard deviation can then be calculated for these combined runs. It should be quantitatively determined if it is acceptable to combine runs. Sometimes particular runs should not be combined with the others because they are not similar enough. The Z value calculation detailed in Section 5.2.2 is used in the following two examples to determine which runs can be combined with 95% confidence.

Example 1:

		Run 1	Run 2	Run 3	Run 4	Run 5	Average
	Average	0.8467	0.8533	0.8240	0.8453	0.852	0.84427
	Std dev	0.0299	0.0324	0.0259	0.0192	0.0286	0.02721
	# points	15	15	15	15	15	75
95% conf int	Minimum	0.0148	0.0160	0.0128	0.0095	0.0141	0.01344
for std dev	Maximum	0.0451	0.0489	0.0389	0.0289	0.0431	0.04097

Table 5.2: Database values for several runs for Example 1

	Z	Z<1.96	Z>-1.96	Combinable?
Run 1 & Run 2 Comparison	-0.5851	1.0	1.0	1.0
Run 2 & Run 3 Comparison	2.7386	1.0	-	-
Run 3 & Run 4 Comparison	-2.5644	-	1.0	-
Run 1 & Run 3 Comparison	2.2199	1.0	-	-
Run 1 & Run 4 Comparison	0.1452	1.0	1.0	1.0
Run 2 & Run 4 Comparison	0.8217	1.0	1.0	1.0
Run 1 & Run 5 Comparison	0.1194	1.0	1.0	1.0
Run 2 & Run 5 Comparison	-0.4992	1.0	1.0	1.0
Run 3 & Run 5 Comparison	-2.8134	-	1.0	-
Run 4 & Run 5 Comparison	-0.7495	1.0	1.0	1.0

Table 5.3: Comparison of runs for Example 1

In Table 5.2, column 2 is equal to 1.0 if $Z<1.96$ and -- if $Z>1.96$. Column 3 is equal to 1.0 if $Z>(-1.96)$ and -- if $Z<(-1.96)$. Column 4 is equal to 1.0 if Columns 2 and 3 are both equal to 1.0. Combinable column (column 4) equals 1.0 if the runs can be combined. Combinable column equals -- if runs cannot be combined. Run 3 cannot be combined with runs 1, 2, 4 or 5; therefore, run 3 should not be included in any averages of the data.

Example 2 (Tables 5.3 and 5.4) shows that a run should be eliminated if it cannot be combined with the majority of other runs (at least 75%).

Example 2:

		Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Average
	Average	0.538	0.554	0.569	0.527	0.562	0.56	0.552
	Std dev	0.0545	0.0408	0.032	0.042	0.0291	0.042	0.040
	# points	25	25	25	25	25	25	150
95% conf int	Minimum	0.0332	0.0248	0.020	0.025	0.018	0.026	0.024
For std dev	Maximum	0.0760	0.0569	0.045	0.058	0.041	0.059	0.056

Table 5.4: Database values for several runs for Example 2

	Z	Z<1.96	Z>-1.96	Combinable?
Run 1 & Run 2 Comparison	-1.1741	1.0	1.0	1.0
Run 2 & Run 3 Comparison	-1.4705	1.0	1.0	1.0
Run 3 & Run 4 Comparison	4.0510	1.0	-	-
Run 1 & Run 3 Comparison	-2.4698	-	1.0	-
Run 1 & Run 4 Comparison	0.8267	1.0	1.0	1.0
Run 2 & Run 4 Comparison	2.3464	1.0	-	-
Run 1 & Run 5 Comparison	-0.7977	1.0	1.0	1.0
Run 2 & Run 5 Comparison	-1.9416	1.0	1.0	1.0
Run 3 & Run 5 Comparison	0.8422	1.0	1.0	1.0
Run 4 & Run 5 Comparison	-3.4852	-	1.0	-
Run 1 & Run 6 Comparison	-1.59802	1.0	1.0	1.0
Run 2 & Run 6 Comparison	-0.51204	1.0	1.0	1.0
Run 3 & Run 6 Comparison	0.8798	1.0	1.0	1.0
Run 4 & Run 6 Comparison	-2.8221	-	1.0	-
Run 5 & Run 6 Comparison	0.19575	1.0	1.0	1.0

Table 5.5: Comparison of runs for Example 2

In example 2, run 4 cannot be combined with runs 2, 3, or 5; therefore, run 4 should not be included in any averages of the data, even though it could be combined with run 1. Run 3 cannot be combined with runs 1 or 4, but run 4 has already been excluded from the combination. Since run 3 can be combined with most (at least 75%) of the runs after run 4 is excluded, it can be included in the combination of the data.

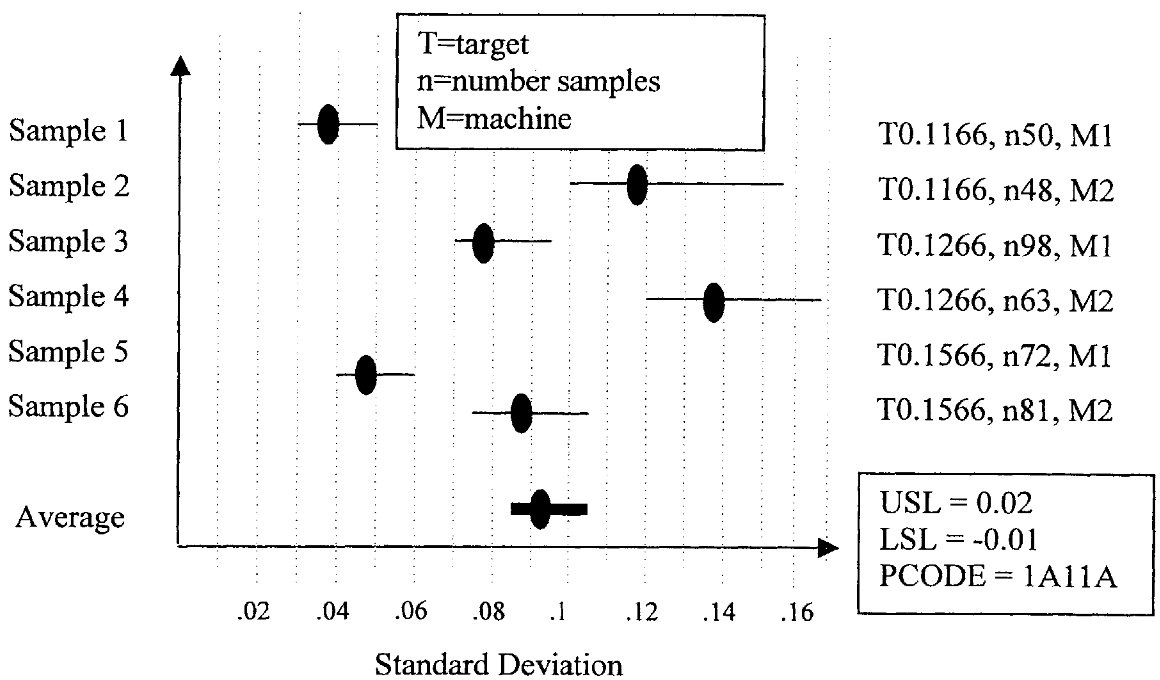
5.4 Aggregate data or numerous samples

There should be a method to obtain statistically valid data when all three components of the index information (material, feature, and process) are not known. "Aggregate data" is the term used for the data when either all three parameters of material, feature, and process are not known or when all the details of one or more of these three parameters are not known. Providing aggregate data allows PCD to be used to select between sub-processes for a design. Section 5.4.1 explains how this aggregate data can be plotted as a group of samples. Determining which outlier samples should be excluded from averages for the aggregate data is explained in Section 5.4.2. Finally, Section 5.4.3 details how groupings of samples can be combined.

5.4.1 Plotting aggregate data

If a designer knows he/she wants to make a hole by drilling with a desired tolerance of 0.002 inches, he/she may want to look at the data for various materials to determine which should be best for this tolerance for this particular feature and process. Alternatively, the designer may want to look at the average data for all or several materials. Aggregate data should also be available when one doesn't know all the details of particular material, feature, or process component, but does know some of the details. For example, the designer may know he/she wants to use a forging but not know if he/she want to use hand or die forging.

How should this aggregate data be provided to the designer? If all the data meeting the specified criteria is simply averaged, then a high level of uncertainty in the data is probable. It is likely there are problems such as data outliers, unreliable data, and/or data groupings, which will cause this average value to be of minimal validity. An alternative to simply providing the average value is to present all the detailed data for the samples that meet the selections the designer has made. For example, if the designer knows he/she want to drill a one inch hole, but he/she does not know which material to use, then a plot of the various material options should be provided. Figure 5.8 shows what this situation might look like in the fraemwork developed in previous sections of this thesis. Again this is fictitious data. The confidence interval for each sample is based on the combination and exclusion of runs detailed in Sections 5.3.2 and 5.3.3.



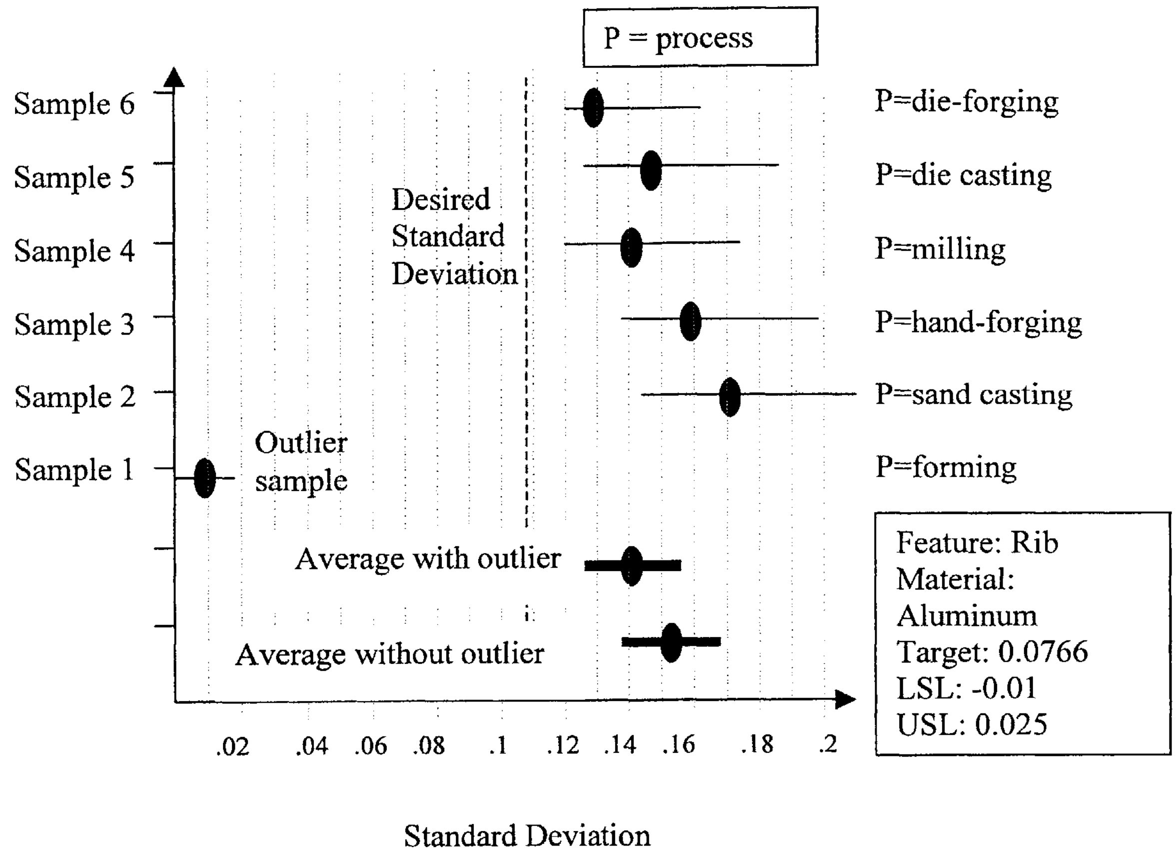
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Total
Std dev	0.04	0.12	0.08	0.14	0.05	0.09	
No. points	50	48	98	63	72	81	412
95% conf int min	0.03341	0.09989	0.07015	0.11911	0.043	0.077955	
95% conf int max	0.04985	0.15033	0.09309	0.16986	0.0598	0.106486	
conf int length	0.01644	0.05044	0.02294	0.05075	0.0169	0.028531	
average std dev							0.091176
95% conf int min							0.084509
95% conf int max							0.098995
conf int length							0.014485

Figure 5.8: Standard deviation range visualization for multiple samples

In Figure 5.8, since the data is for varying samples, it is expected that the standard deviation values may be significantly different. A similar aggregate data visualization to that shown in Figure 5.8 has been implemented in the software prototype as shown in Section 5.5.3 and 5.5.4.

5.4.2 Excluding aggregate data

When investigating a set of samples for aggregate data, an outlier may occur. In Figure 5.9, fictitious samples are plotted for the various processes that could be used to manufacture an aluminum rib.



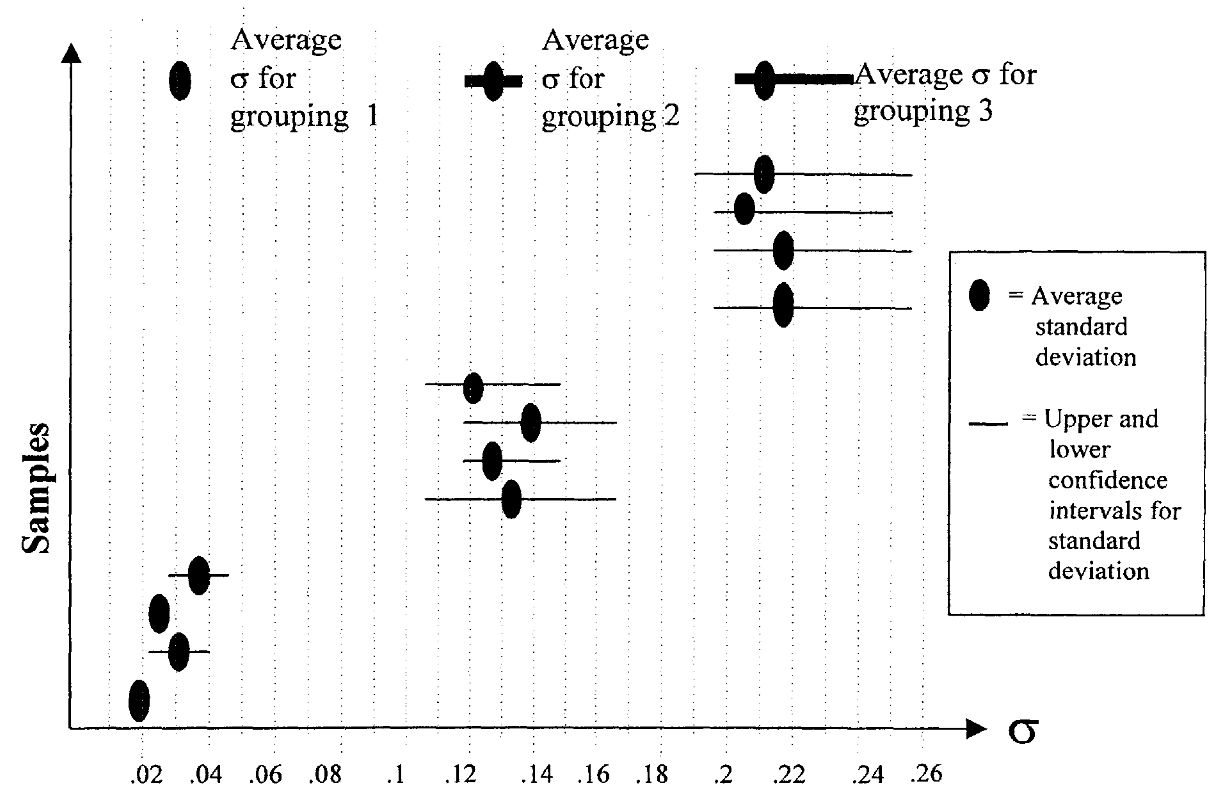
SET	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Total
Std dev	0.012	0.17	0.16	0.14	0.15	0.13	
No. points	42	61	53	45	51	57	309
95% conf int min	0.00987	0.14428	0.13429	0.1159	0.1255	0.10975	
95% conf int max	0.01531	0.20698	0.19799	0.17688	0.1865	0.1595	
conf int length	0.00543	0.06271	0.0637	0.06098	0.061	0.04975	
average std dev w. outlier							0.13963
95% conf int min w. outlier							0.12942
95% conf int max w. outlier							0.151602
conf int length							0.022183
average std dev w/o outlier							0.150175
95% conf int min w/o outlier							0.139178
95% conf int max w/o outlier							0.163075
conf int length w/o outlier							0.023898

Figure 5.9: Depiction of outlier sample for a set

In Figure 5.9, Sample 1 (forming) is an outlier because it over 10% times less than the average of all the samples in the set. Therefore, forming should be excluded in the average for this aggregate data for aluminum ribs.

5.4.3 Grouping aggregate data

In plotting aggregate data, groupings in the data are more likely to occur than single sample outliers. These groupings occur because of the high variability in the samples for aggregate data. An example of this is shown in Figure 5.10 for a group of fictitious data. There are three distinct groupings of data each with its own average standard deviation. The data within each individual grouping could be combined, but the data from the three separate groupings should not be combined. The combination of data within groupings should be determined using the Z value discussed in Section 5.2.3.



	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	S 9	S 10	S 11	S 12
Std dev	0.02	0.03	0.025	0.035	0.13	0.125	0.135	0.12	0.22	0.215	0.205	0.21
No. points	45	57	68	72	47	92	71	83	93	82	62	70
95% conf int min	0.017	0.025	0.021	0.030	0.108	0.110	0.116	0.104	0.192	0.186	0.174	0.181
95% conf int max	0.025	0.037	0.030	0.042	0.163	0.146	0.162	0.142	0.257	0.254	0.249	0.252
conf int length	0.009	0.011	0.009	0.012	0.055	0.031	0.046	0.038	0.065	0.068	0.075	0.072

	Grouping 1	Grouping 2	Grouping 3
average std dev	0.02858	0.126272	0.2124
95% conf int min	0.02624	0.116808	0.1938
95% conf int max	0.03138	0.137418	0.2349
conf int length	0.00514	0.02061	0.0411

Figure 5.10: Breakdown of aggregate data to determine groupings

5.5 Prototype software

This section describes the prototype software. First, Section 5.5.1 provides a description of the database of sample data used in the prototype software. Next, the methodology of the software is described in Section 5.5.2. The prototype software plots either the mean shift or the standard deviation with its upper and lower limits for a 95% confidence interval in order to add clear visualization to the data. The inclusion of uncertainty and the elimination of outlier data in the prototype software add statistical validity to the data. Sections 5.5.3 and 5.5.4 provide further details on the two Visual Basic forms that serve as the user interface and output, respectively. Section 6.2 explains how the software determines outlier data to exclude. Finally, examples of using the prototype software for plotting standard deviation and mean shift are provided in Sections 6.3 and 6.4.

5.5.1 Database structure

The database used in the prototype software was provided by the large aerospace company; however, the data has been disguised. This disguising of the data is probably the reason for the peculiar target values of 0.0766, 0.1016, etc. rather than 0.75 and 0.1, etc. It may also be the cause for some of the lower specification values being positive. Since LSL is defined as the difference between the lower tolerance and the target, it should always be negative. The lower tolerance value should always be smaller than the target value. The process capability data the large aerospace company collects and records includes the target dimension value and the upper and lower specification limits for this target, the index, the machine used, and the dimension obtained (measurement). The index encompasses the process, material, and feature.

5.5.2 Software implementation steps

Figure 5.11 shows the methodology of the prototype software. First the options for the sorting parameters are presented to the user, then the data is sorted in ACCESS, and finally the data is plotted in Visual Basic. For each step in Figure 5.11, it is indicated if the step is performed in ACCESS or Visual Basic. The two Visual Basic forms are also indicated in this figure.

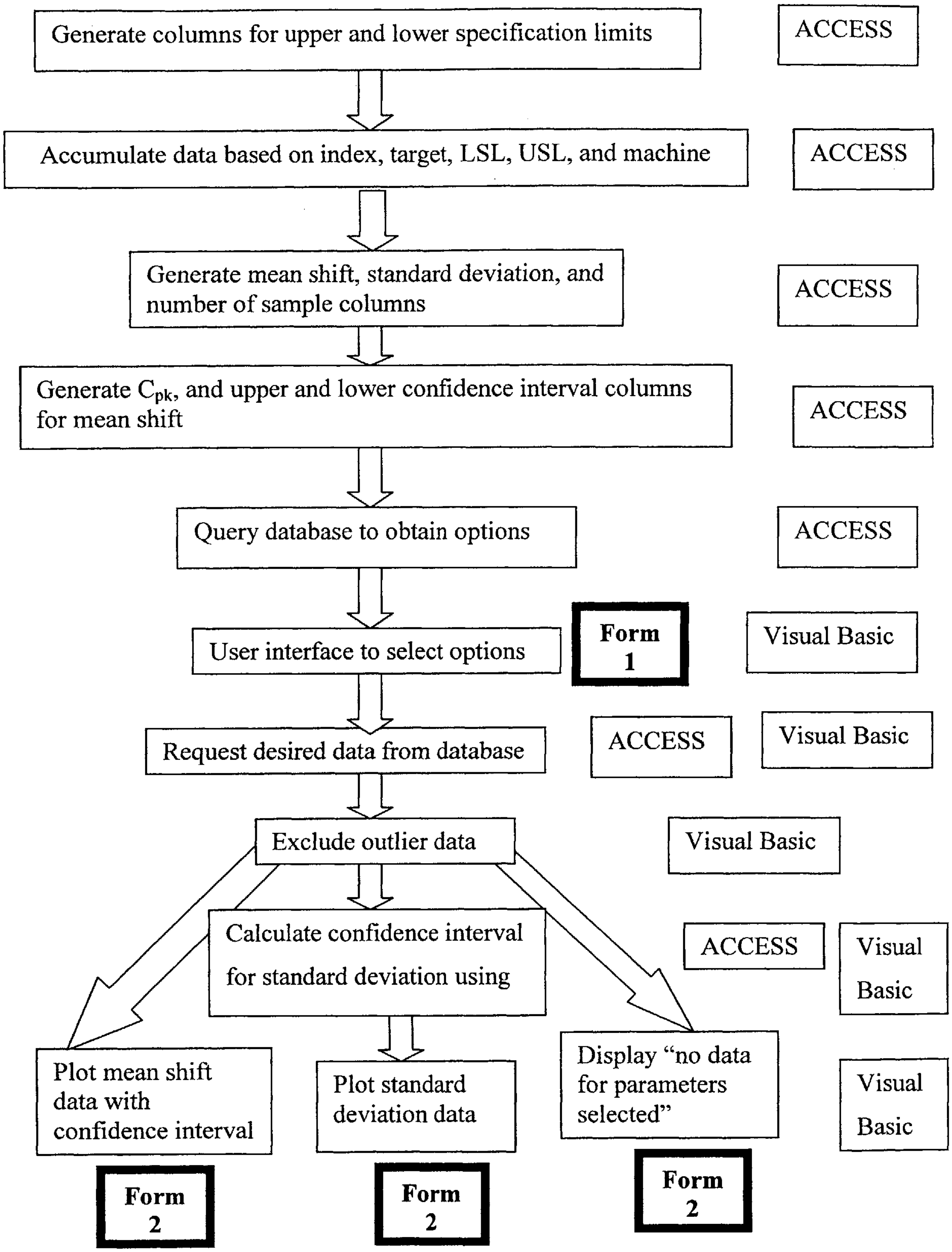


Figure 5.11: Software methodology steps and programs used

In Figure 5.11, the box labeled "User interface to select options" is form 1 and this is where the designer would chose the options he/she would like to sort the data by. This is equivalent to where the designer would have to progress through the hierarchy to find the desired index in Figure 4.1. One difference between Figure 4.1 and the prototype software, is that the user inputs either the desired mean shift or the desired standard deviation rather than the desired tolerance; however, the desired tolerance value could simply be substituted in place of these values.

Next, Visual Basic determines and excludes outlier data. The prototype software does not calculate the Z-value; therefore, it does not determine if runs can be combined and it does not determine groupings of samples. Form 2 plots the data to allow the user to visualize which runs fall below the target mean shift or standard deviation value. More details on forms 1 and 2 are provided in Sections 5.5.3 and 5.5.4.

5.5.3 User interface form

An example of the user interface form for mean shift is shown in Figure 5.12 and for standard deviation is shown in Figure 5.13. Each box in this form is labeled with a number. When the boxes are described, the number of the box is provided. The first Visual Basic form allows the user to select a single value for each of the following parameters: PCODE (1), minimum LSL (11), maximum USL (12), minimum target (9), and maximum target (10). The form also allows the user to select any number of machines (7). The user can input a value for the "maximum number of lines to plot" (8) and the "number of intervals on x-axis" (4). The default value for both the maximum number of lines and for the number of intervals on x-axis is 10.

Form 1 also allows the user to chose whether they want to plot standard deviation or mean shift with a 95% confidence interval (6). Depending on which of these is chosen, the user can input a value for the "desired maximum mean shift" (2) or for the "desired standard deviation" (3). The default value for the desired mean shift is 0.005 and for the standard deviation is 0.01. If neither of the mean shift or standard deviation options is chosen, the default is standard deviation. The explanation of the PCODE, which explains what each digit of the index means is proprietary but is provided in the software prototype (5).

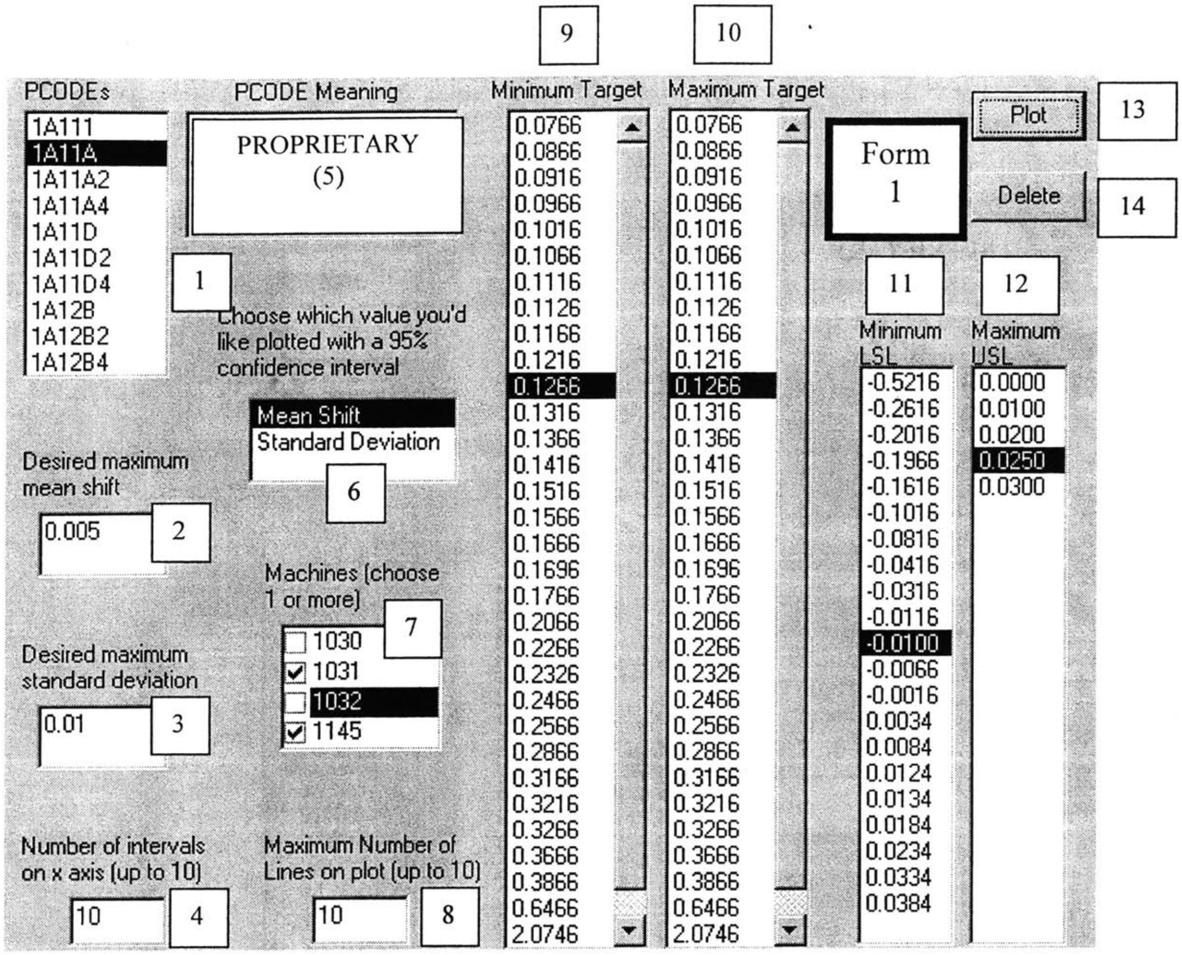


Figure 5.12: Sample of user interface form for mean shift selection

The options presented to the user on form 1 are generated directly from the ACCESS database. The program looks for all the options of values for the index, machine, target value, LSL, and USL that are available in the ACCESS database. This greatly reduces the probability the user will chose an index that is not populated or is infeasible because the user can choose only those values available in the data set.

The program ensures no value appears more than once as an option. Once the lists of choices for each parameter are listed in Visual Basic, the user chooses one value for each parameter (PCODE, LSL, USL, minimum target, and maximum target). The user can choose more than one machine. If a value is not chosen for either the machine or index parameter then the entire range of values for that parameter is used. For example, if the user does not chose a desired index, then the data is sorted for the machines selected, the minimum and maximum target selected, the LSL selected, the USL selected, and for all indexes.

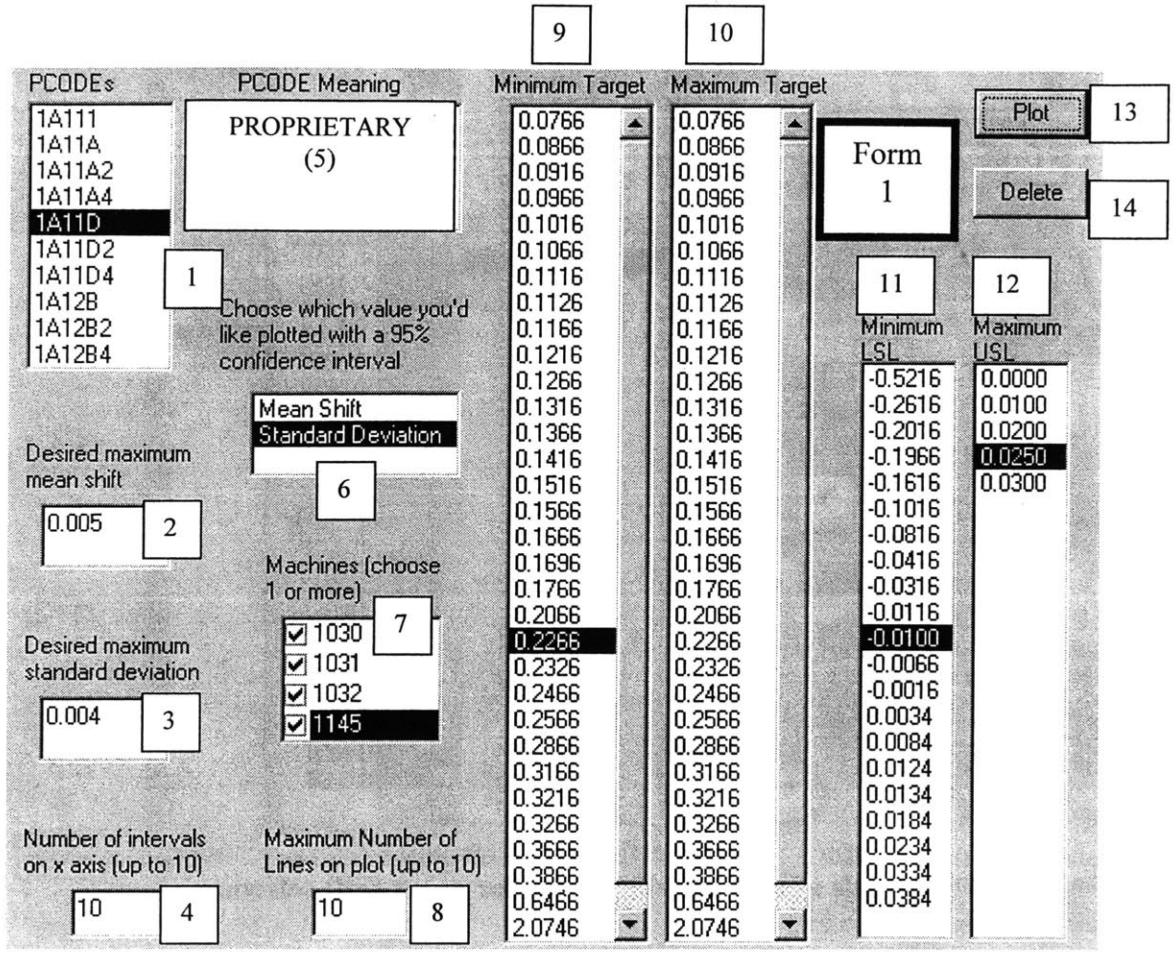


Figure 5.13: Sample of user interface form for standard deviation selection

The user must select a minimum or a maximum target, but does not need to select both unless a range of target values is desired. If no value is selected for either USL or LSL, then the entire range of values between the minimum LSL and zero and between the maximum USL and zero is used. When the user selects a desired index on form 1, the meaning of that index (called a PCODE for the company studied) is displayed beneath it.

After the user has selected one of each parameter (or more than one for the machines) and completed the other user-input values, he/she selects the "plot" command (13). This brings the user to form 2. The "delete" command button (14) is used to delete the various special databases the program creates in ACCESS for each of the parameters. This should be used whenever there is a problem and the index produces a run error. If this "delete" button is not pressed, the user must go into ACCESS and manually deleted the special databases. If the program runs without any errors, the delete command does not need to be pressed because the special databases will be deleted automatically.

Once the user presses the "plot" command button on form 1, the options chosen by the user are used to write some code in Structured Query Language (SQL). This code sorts the database for the desired parameters. All the data for the options selected by the user are placed into a new database called "qdeftest". Then it is determined which data should be excluded.

5.5.4 Data output form

An example of the data output form for mean shift is shown in Figure 5.14 and for standard deviation is shown in Figure 5.15. Each of the features of this form are numbered and are referenced by these numbers when they are described. The output form in Figure 5.14 corresponds to the inputs on the user input form in Figure 5.12. The output form in Figure 5.15 corresponds to the inputs on the user input form in Figure 5.13.

The second Visual Basic form, the data output form, produces a plot and a table of the data when the user presses the "plot" command button (16). The plot has target value as the y-axis (14) and either mean shift range or standard deviation range as the x-axis (13). For the y-axis, each run is provided with its own unique line. Runs with larger target values are higher on the y-axis. However, if there are several runs with the same target value, they will be at different heights so that they will each have their own line.

For each sample of data a line is plotted (11). The circle at the center of this line represents either the mean shift or the standard deviation value (12). The line runs from the lower to the upper confidence interval value for this mean shift for a 95% confidence. The "clear" command button is used to clear the plot (17).

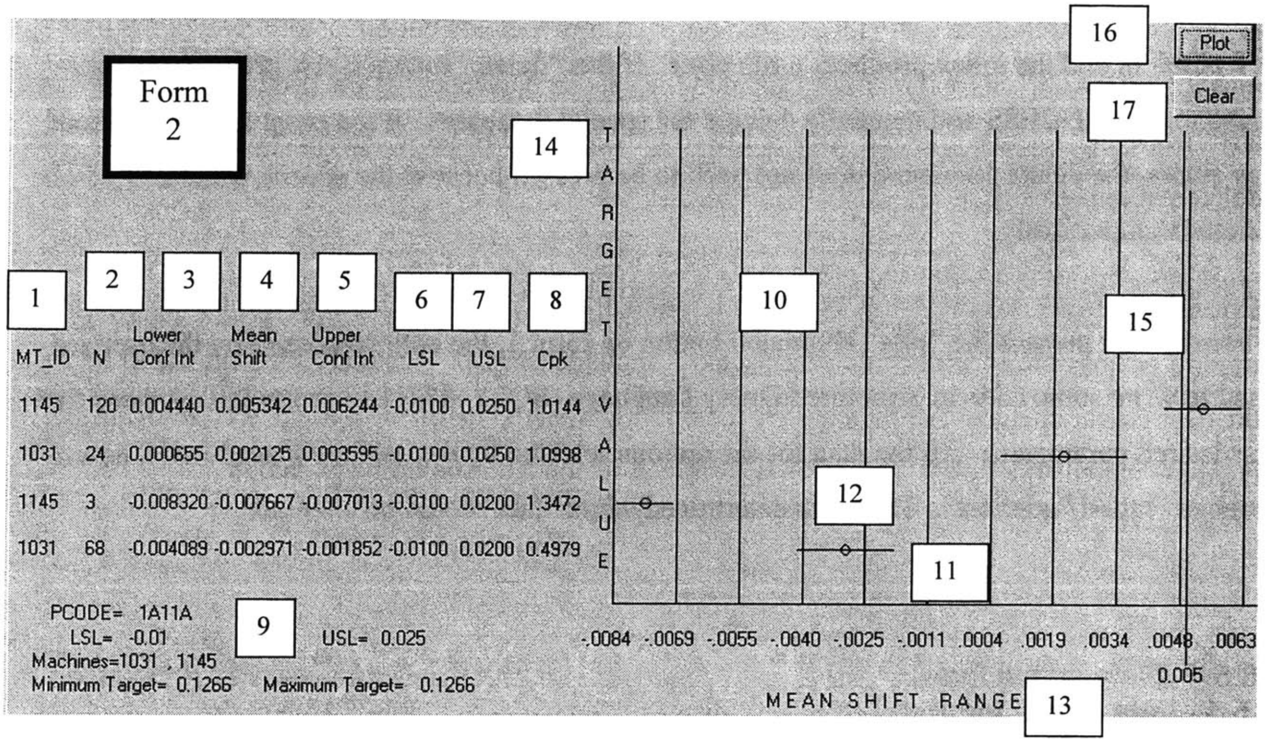


Figure 5.14: Sample of output form for mean shift selection

The software prototype determines the minimum lower confidence interval and the maximum upper confidence interval for all the samples and these become the range for the x-axis. This zooms the plot. Then, equally spaced intervals (10) along the x-axis are created based upon either the default value or the user input value for the "number of intervals on x-axis".

For the "desired maximum mean shift" and "desired standard deviation" values (15), if the default or user input value is between the minimum and maximum range plotted, then this value will be printed as a vertical line with a label of what the value is. If the desired value is less than the range plotted, then the desired value is not plotted and a message is printed below the indicating whether the desired value is above or below the range.

A table is created to the left of the plot on form 2 (1-8). This table shows the values for each line that is plotted. The data for each plotted line is written in the horizontal line in the table corresponding to the plotted line. The data in the table includes: the machine (1), the number of samples (2), the lower confidence interval value (3), the mean shift or standard deviation (whichever was chosen) (4), the upper confidence interval value (5), the lower specification limit (6), the upper specification limit (7) and the C_{pk} value (8). Below the table all of the values for the parameters the user selected are listed (9). Chapter 6 show several examples of what the output form of Visual Basic looks like for various inputs.

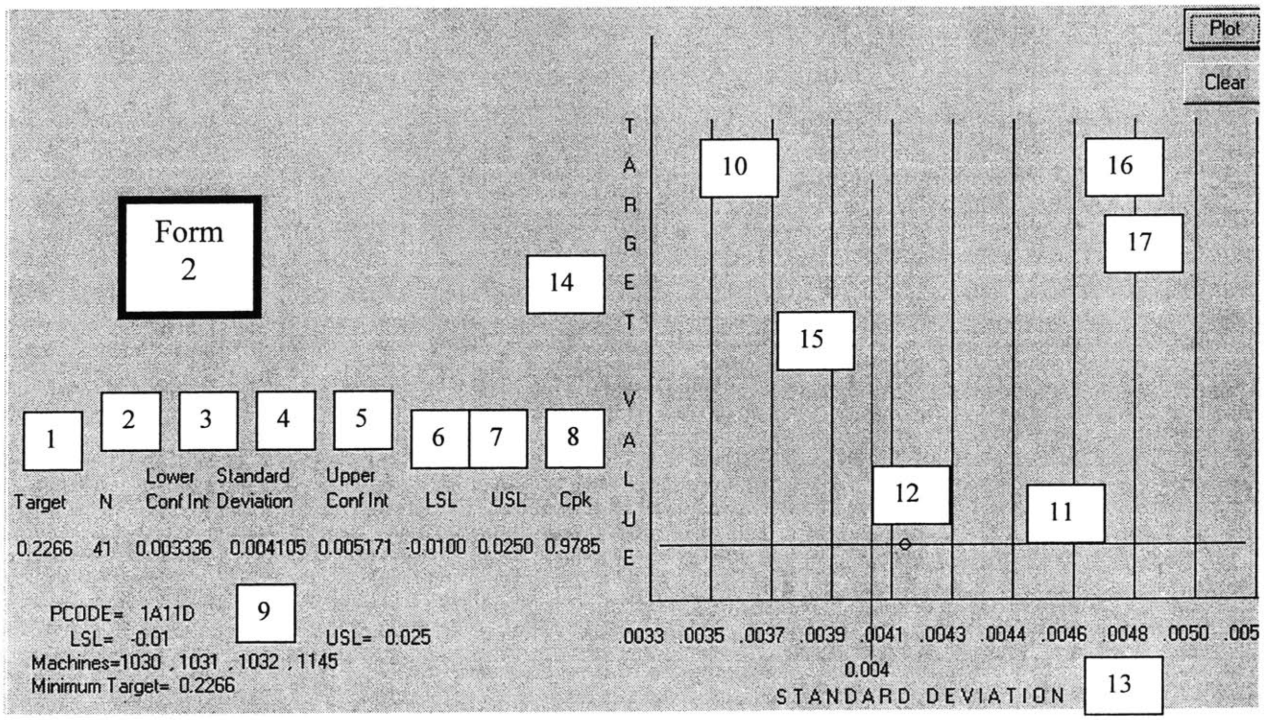


Figure 5.15: Sample of user output for standard deviation selection

5.6 Conclusion

This chapter addresses five of the technical barrier to PCD usage by design. First, this chapter provided the theory for quantifying uncertainty. Second, it provided quantitative methods to determine if runs/samples can be combined and visual methods to determine outlier data. Third, a means of providing appropriate aggregate data and determining groups in it was detailed. Fourth, the prototype software system presents the PCD graphically and also shows how the uncertainty can be portrayed as a confidence interval. The prototype software can easily be modified to address the fifth issue of not allowing infeasible indexes to be chosen. The last two technical barriers to PCD usage by design are addressed in Chapter 7.

6 Features of Software Prototype

6.1 Introduction

This chapter provides examples of the special features of the prototype software. Among these features are the ability to choose multiple values for the sorting parameters and the automatic exclusion of outlier data. Examples are provided for plotting both the mean shifts and the standard deviations.

6.2 Exclusion of outlier data

In order to augment the statistical validity of the data and to allow for clarity in the data plots, outlier data was excluded in the prototype software. If outlier data had not been excluded than the plots would have been unclear because the outlier data would determine the minimum or maximum value of the x-axis. The outlier would greatly expand the range of values along the x-axis; therefore, the other plotted lines would be much smaller because the x-axis intervals would span such a large range of values.

Many exclusion tests were used in the prototype software and for each of them the entire line of data is excluded from both the plot and the table. The first three tests determine if the confidence intervals are out of range. There is a dichotomy of values in the PCODEs in the data provided by the large aerospace company. For the PCODEs, 1A111, 1A11A, 1A11A2, 1A11A4, 1A11D, 1A11D2, and 1A11D4 the lower and upper confidence interval values for mean shift predominantly range between –0.008 and 0.0075. For the PCODEs 1A12B, 1A12B2, and 1A12B4 the mean shift lower and upper confidence interval values predominantly range between 0.012 and 0.028.

There are separate tests for the two PCODE types. For the PCODES 1A111, 1A11A, 1A11A2, 1A11A4, 1A11D, 1A11D2, and 1A11D4, the first test is to see if the upper specification limit is greater than 0.0099. The second test is to see if the lower specification limit is less than –0.0099. These values were determined based on looking at all the data in the database provided by the large aerospace company and determining that most of the data for these PCODEs fell between –0.005 and 0.009. Then values beyond the values outside the –0.005 to 0.009 range were excluded. The average for all the runs for these PCODEs for the upper confidence interval was 0.001334 with 4193 points in 1200 runs. After excluding the 15 outlier runs, the average upper confidence interval was 0.001105 with 3998 points. Values greater than 0.0099 are outliers because they are significantly over 3 times more than the average value. The average for all the runs for these PCODEs for the lower confidence interval was (-0.000469) with 4193 points in 1200 runs. After excluding the one outlier run, the average lower confidence interval was (-0.000394) with 4154 points. Values less than (-0.0099) are outliers because they are over 3 times less than the average value.

These outliers for the upper and lower confidence interval for the mean shift should not be included because they are likely errors. Also, plotting these values greatly unzooms the plot because these values are so much higher than the rest that all of the other lines become simply points.

For the PCODEs 1A12B, 1A12B2, and 1A12B4, the first test is to see if the upper specification limit is greater than 0.03. The second test determines if the lower specification limit is greater than 0.03. These values were determined based on looking at all the data in the database provided by the large aerospace company and determining that most of the data for these PCODEs fell between 0.0001 and 0.03. Then values beyond the values outside the 0.001 to 0.03 range were excluded. The average for all the runs for these PCODEs for the upper confidence interval was 0.00105 with 1301 points in 46 runs. After excluding the 3 outlier runs, the average upper confidence interval was 0.000755 with 1280 points. Values greater than 0.03 are outliers because they are significantly over 3 times more than the average value. The average for all the runs for these PCODEs for the lower confidence interval was 0.00055 with 1301 points in 46 runs. After excluding the one outlier run, the average lower confidence interval was 0.0006 with 1292 points. Values less than 0.0001 are outliers because they are over 3 times less than the average value.

There is not a dichotomy between the PCODEs for standard deviations. The confidence interval values for all PCODEs predominantly fall between 0.0 and 0.015. The third test is to see if the standard deviation is greater than 0.015, which would mean the run is an outlier. This third test is only performed if the user chooses to plot standard deviation.

For all PCODEs, the fourth test is to determine if C_{pk} is greater than 5.0 or if C_{pk} is less than (-5.0). The fifth test is to determine if the number of samples is less than 2. The final test is to see if the number of lines of data, after the prior four exclusions, is greater than the user input or default value for the maximum number of lines to plot. After all the exclusion tests are performed, the table and plot are generated on the second Visual Basic form once the user hits the "plot" command button. If there is no data to plot either before or after the exclusion process, then the program simply prints on the plot screen, that there is no data available.

All exclusion tests are performed regardless of whether the user chooses to plot mean shift or standard deviation.

6.3 Mean shift examples

The various software examples that follow show the various capabilities of the software for plotting mean shift. Some of the capabilities, such as sorting by multiple values for the parameters of PCODE, are not current capabilities, but instead capabilities that can easily be added to the program. The features included are the listing of possible values for PCODE, minimum target, maximum target, machine, LSL, and USL. The meaning of each PCODE is also displayed. The user can change the default values for "desired maximum mean shift", "desired maximum standard deviation", "maximum number of lines on plot", and "number of intervals on x-axis".

In the program, it is easier to distinguish between the various lines of data because a random color generator creates different colors for each line. Also, the "desired maximum mean shift" or "desired standard deviation" value is plotted as a vertical line in red. The colors had to be changed to only darker colors so they could be seen in black and white.

There are eight examples of using the prototype software to plot the mean shift. The first example shows how the software plots the data only for the machines chosen. It also shows how the software plots the user-input desired maximum mean shift. The second example shows that the user can input a maximum number of lines to plot. It also shows what the prototype software does when the user-input desired maximum mean shift is outside the range of the mean shift plotted. The third example shows a comparison of various PCODEs that begin with 1A11A. The fourth example shows a comparison of various PCODEs that begin with 1A12B. The fifth example shows how the software excludes the data when the confidence interval for the mean shift is an outlier. The sixth example shows that the user can input the number of intervals to plot in the output of the software. The seventh example shows how the prototype software allows the user to plot a range of target values. The eighth example shows the output of the software when there is no data available for the parameters selected by the user.

For each example, most of the input parameters are listed in the Table on form 2. The values for the other parameters are provided before the plot.

<u>Form 1 Data Input</u>

Plot: Mean shift

Desired maximum mean shift: 0.005 (DEFAULT)

Number of intervals on x-axis: 10 (DEFAULT)

Maximum number of lines on plot: 10 (DEFAULT)

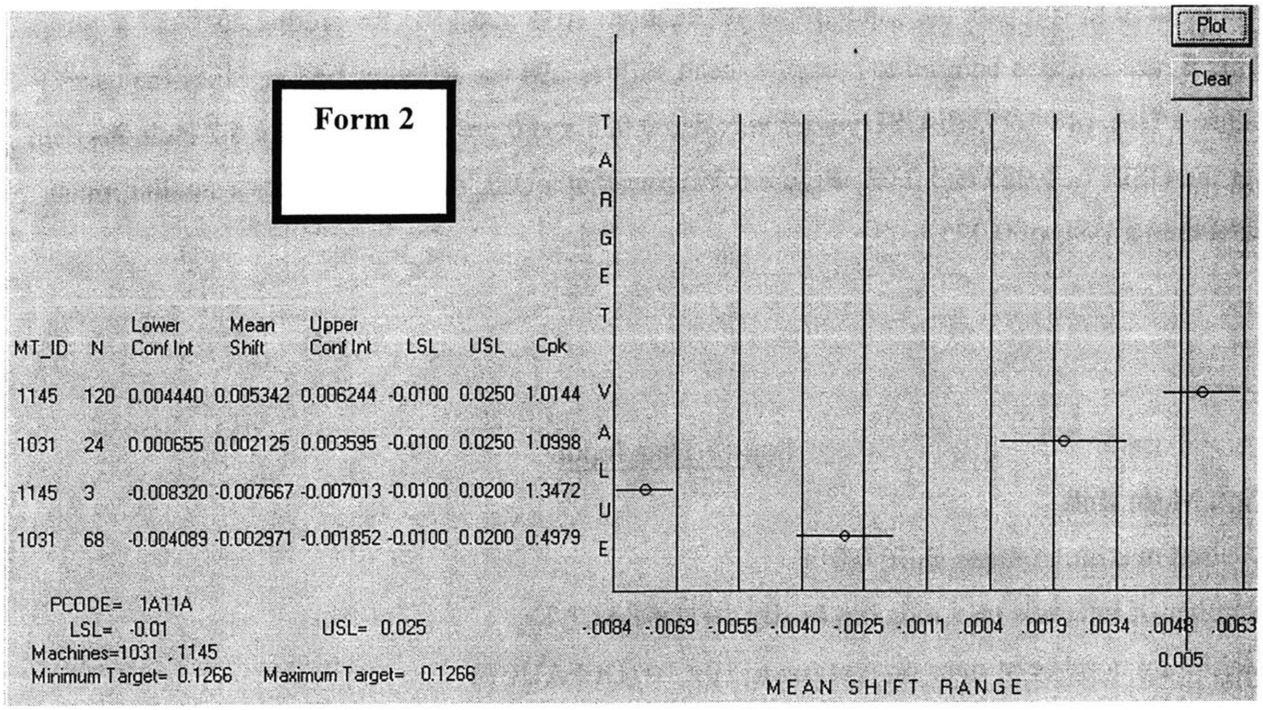


Figure 6.1 Prototype software screen for PCODE 1A11A

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev	Cpk	Lower Conf	Upper Conf
1A11A	0.12660	-0.0100	0.0200	1031	68	-0.0029706	0.005	0.4979	-0.00408918	-0.001852
1A11A	0.12660	-0.0100	0.0200	1145	3	-0.0076667	0.0006	1.3472	-0.00832000	-0.0070133
1A11A	0.12660	-0.0100	0.0250	1031	24	0.00212500	0.004	1.0998	0.000654704	0.0035953
1A11A	0.12660	-0.0100	0.0250	1032	108	0.0025	0.012	0.3466	0.000232458	0.00476754
1A11A	0.12660	-0.0100	0.0250	1145	120	0.00534167	0.005	1.0144	0.004439643	0.00624369

Figure 6.2: Data for Figure 6.1

The sample of data in Figure 6.1 shows how the prototype software plots the data for the options selected for each parameter. Figure 6.2 shows there are five lines of data; however, only machines 1031 and 1145 were selected. Therefore, line 4 of Figure 6.2 is not included in the plot, since it is for machine 1032.

The user did not change the default value of 0.005 for the desired maximum mean shift; therefore, this value is plotted as a vertical line, which is labeled. The data plotted in Figure 6.1

can be used to compare the capabilities of machines 1031 and 1145 for producing 1A11A parts. This figure can also be used to compare mean shift results for different USLs. Since the user chose a USL of 0.025, all USL values between 0.025 and 0 are plotted. Figure 6.2 includes data for the USLs of 0.025 and 0.02. Figure 6.1 shows that a USL of 0.02 produces a smaller mean shift than a USL of 0.025.

Form 1 Data Input

Plot: Mean shift

Desired maximum mean shift: 0.011

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

PCODE	N	Lower Conf Int	Mean Shift	Upper Conf Int	LSL	USL	Cpk
1A11D4	57	-0.000283	0.000544	0.001371	-0.0100	0.0200	1.1034
1A11D2	9	0.002271	0.004000	0.005729	-0.0100	0.0200	1.7638
1A11D2	60	0.000829	0.001667	0.002505	-0.0100	0.0200	1.1741
1A11D	30	0.000621	0.001300	0.001979	-0.0100	0.0200	1.9862
1A11D	419	-0.000181	0.000169	0.000520	-0.0100	0.0200	0.9270
1A11A4	24	0.000754	0.002500	0.004246	-0.0100	0.0200	0.9548
1A11A4	24	-0.001739	-0.000542	0.000656	-0.0100	0.0200	1.0536
1A11A2	12	0.004874	0.007750	0.010626	-0.0100	0.0200	0.8033
1A11A	6	-0.002344	-0.001167	0.000011	-0.0100	0.0200	2.0004
1A11A	136	-0.000714	0.000074	0.000861	-0.0100	0.0200	0.7165

PCODE= 1A11A, 1A11A2, 1A11A4, 1A11D, 1A11D2, 1A11D4
LSL= -0.01 USL= 0.02
Machines=1031 , 1032 , 1145
Minimum Target= 0.0766 Maximum Target= 0.0766
Desired maximum mean shift is larger than the plotted mean shift range

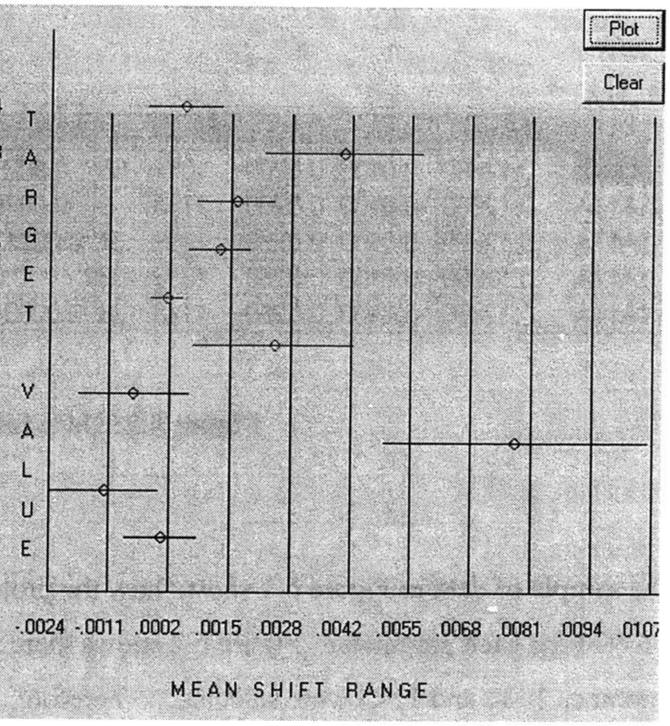


Figure 6.3: Prototype software screens for comparison of 1A11 options

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDevC	Cpk	Lower Conf	Upper Conf
1A11A	0.07660	-0.0100	0.0200	1031	136	7.3529E-05	0.0047	0.7165	-0.0007142	0.0008612
1A11A	0.07660	-0.0100	0.0200	1145	6	-0.0011667	0.0015	2.0004	-0.0023445	1.1147E-0
1A11A2	0.07660	-0.0100	0.0200	1031	12	0.00775	0.0051	0.8033	0.0048738	0.0106262
1A11A4	0.07660	-0.0100	0.0200	1031	24	-0.0005417	0.003	1.0536	-0.0017389	0.0006555
1A11A4	0.07660	-0.0100	0.0200	1145	24	0.0025	0.0044	0.9548	0.00075408	0.0042459
1A11D	0.07660	-0.0100	0.0200	1031	419	0.00016945	0.0037	0.9270	-0.0001807	0.0005195
1A11D	0.07660	-0.0100	0.0200	1145	30	0.0013	0.0019	1.9862	0.00062136	0.0019786
1A11D2	0.07660	-0.0100	0.0200	1031	60	0.00166667	0.0033	1.1741	0.00082852	0.0025048
1A11D2	0.07660	-0.0100	0.0200	1145	9	0.004	0.0026	1.7638	0.00227144	0.0057285
1A11D4	0.07660	-0.0100	0.0200	1031	57	0.00054386	0.0032	1.1034	-0.0002830	0.0013707
1A11D4	0.07660	-0.0100	0.0200	1145	75	-0.0009867	0.003	1.013	-0.0016579	-0.000315

Figure 6.4: Data for Figure 6.3

The sample of data in Figure 6.3 shows what the prototype software does when the user inputs a maximum number of lines to plot that is less than the number of lines of data for the parameters inputted. Figure 6.4 shows there are 11 lines of data for the parameters selected, but Figure 6.3 shows that only the first ten of these are plotted because this is the default maximum number of lines to plot.

This sample also shows that the user input a value of 0.011 for the "desired maximum mean shift" and this value is out of range. In this example, beneath the table, it is indicated that "Desired maximum mean shift is larger than the plotted mean shift range".

The data plotted in Figure 6.3 can be used to compare the capabilities for producing various 1A11 parts. This figure shows that generally a sixth digit of "2" produces a higher standard deviation than a sixth digit of "4" or a blank sixth digit. The ability to choose multiple PCODEs has not yet been incorporated into the software, but this feature can be added.

<u>Form 1 Data Input</u>

Plot: mean shift

Desired maximum mean shift: -0.009

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

Minimum target: -- (NO VALUE SELECTED)

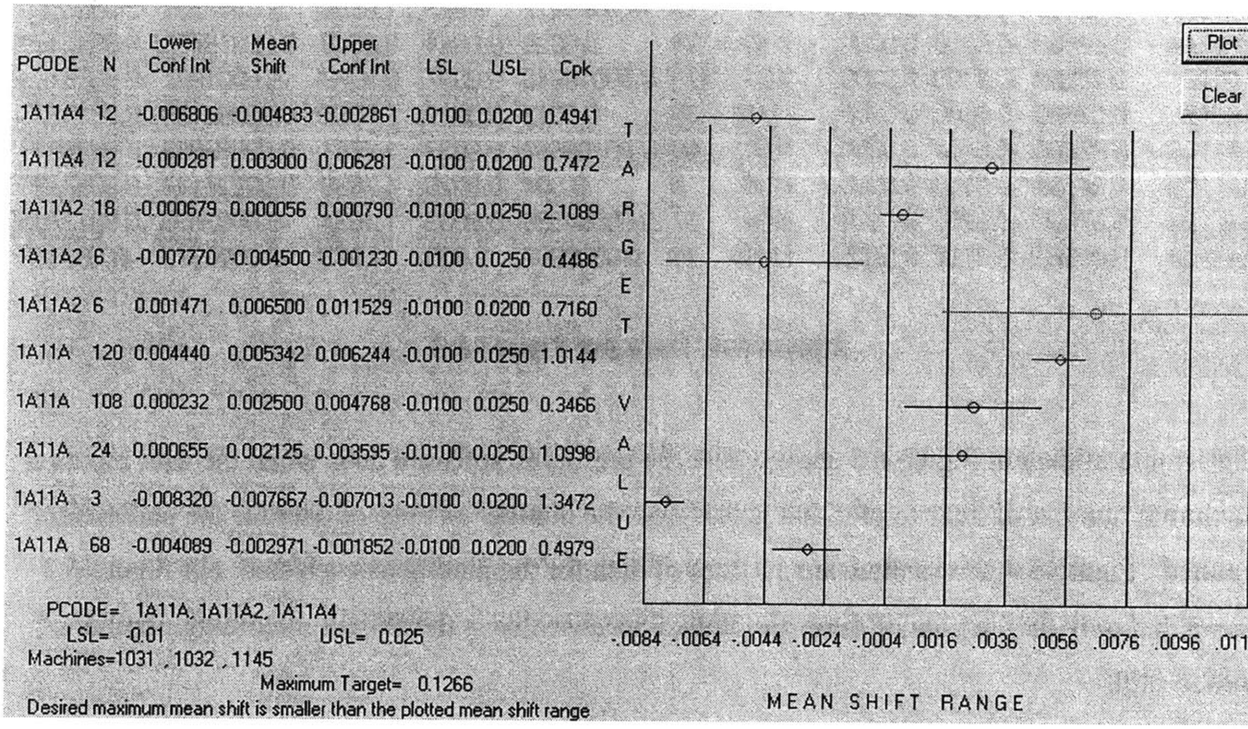


Figure 6.5: Prototype software screen for comparison of 1A11A options

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Conf
1A11A	0.12660	-0.0100	0.0200	1031	68	-0.002971	0.005	0.4979	-0.0040892	-0.001852
1A11A	0.12660	-0.0100	0.0200	1145	3	-0.007667	0.0006	1.3472	-0.0083200	-0.0070133
1A11A	0.12660	-0.0100	0.0250	1031	24	0.0021250	0.004	1.0998	0.00065470	0.0035953
1A11A	0.12660	-0.0100	0.0250	1032	108	0.0025000	0.012	0.3466	0.00023246	0.00476754
1A11A	0.12660	-0.0100	0.0250	1145	120	0.0053417	0.005	1.0144	0.00443964	0.00624369
1A11A2	0.12660	-0.0100	0.0200	1031	6	0.0065000	0.006	0.7160	0.00147103	0.01152897
1A11A2	0.12660	-0.0100	0.0250	1032	6	-0.0045	0.004	0.4486	-0.0077699	-0.0012301
1A11A2	0.12660	-0.0100	0.0250	1145	18	5.556E-05	0.002	2.1089	-0.0006787	0.00078981
1A11A4	0.12660	-0.0100	0.0200	1031	12	0.003	0.006	0.7472	-0.0002815	0.00628148
1A11A4	0.12660	-0.0100	0.0200	1145	12	-0.004833	0.003	0.4941	-0.0068057	-0.002861
1A11A4	0.12660	-0.0100	0.0250	1032	24	-0.003125	0.006	0.4028	-0.0054014	-0.0008486
1A11A4	0.12660	-0.0100	0.0250	1145	78	-0.000103	0.006	0.5796	-0.0013659	0.00116076

Figure 6.6: Data for Figure 6.5

The sample of data in Figure 6.5 allows for a comparison of the PCD for the various PCODEs beginning with 1A11A. The ability to compare various PCODEs is not yet a function of the prototype software; therefore, it was manually inputted into the program to obtain data for these specific PCODEs.

In this figure, the user inputted a desired value for the maximum mean shift that is less than the range of the plot, as indicated in the figure. Figure 6.6 contains 12 lines of data; however, only 10 are plotted in Figure 6.5 because this is the default value for the maximum number of lines. Figure 6.5 also shows that the user only needs to chose either a minimum target or a maximum target value if only one target value, rather than a range of target values, is desired.

Form 1 Data Input

Plot: Mean shift

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

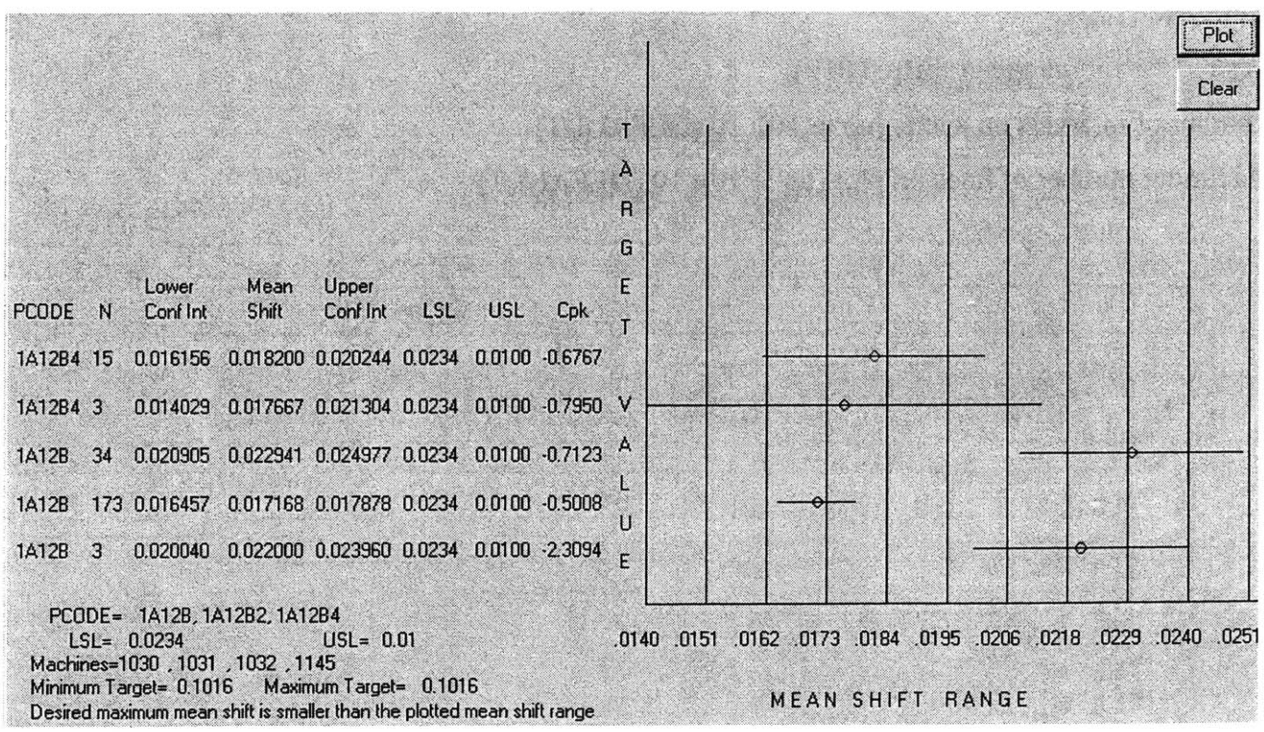


Figure 6.7: Prototype software screen for comparison of 1A12B options

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev	Cpk	Lower Conf	Upper Conf
1A12B	0.10160	0.0234	0.0100	1031	3	0.022	0.002	-2.30940108	0.02004	0.02396000
1A12B	0.10160	0.0234	0.0100	1032	173	0.0171676	0.005	-0.50084166	0.01645677	0.01787849
1A12B	0.10160	0.0234	0.0100	1145	34	0.0229412	0.006	-0.71227579	0.02090544	0.02497691
1A12B2	0.10160	0.0234	0.0100	1145	3	0.02	2E-09	-1549812.83	0.02	0.02000000
1A12B4	0.10160	0.0234	0.0100	1032	3	0.0176667	0.003	-0.7949963	0.01402906	0.02130427
1A12B4	0.10160	0.0234	0.0100	1145	15	0.0182	0.004	-0.67671930	0.01615594	0.02024406

Figure 6.8: Data for Figure 6.7

Figure 6.7 shows how the prototype software could be enhanced to allow for a comparison of the PCODEs beginning with 1A12B. Figure 6.8 shows that the line 4 is not included in the plot because of the exclusion of data with $C_{pk} < -5.0$.

<u>Form 1 Data Input</u>

Plot: Mean shift

Desired maximum mean shift: 0.0191

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

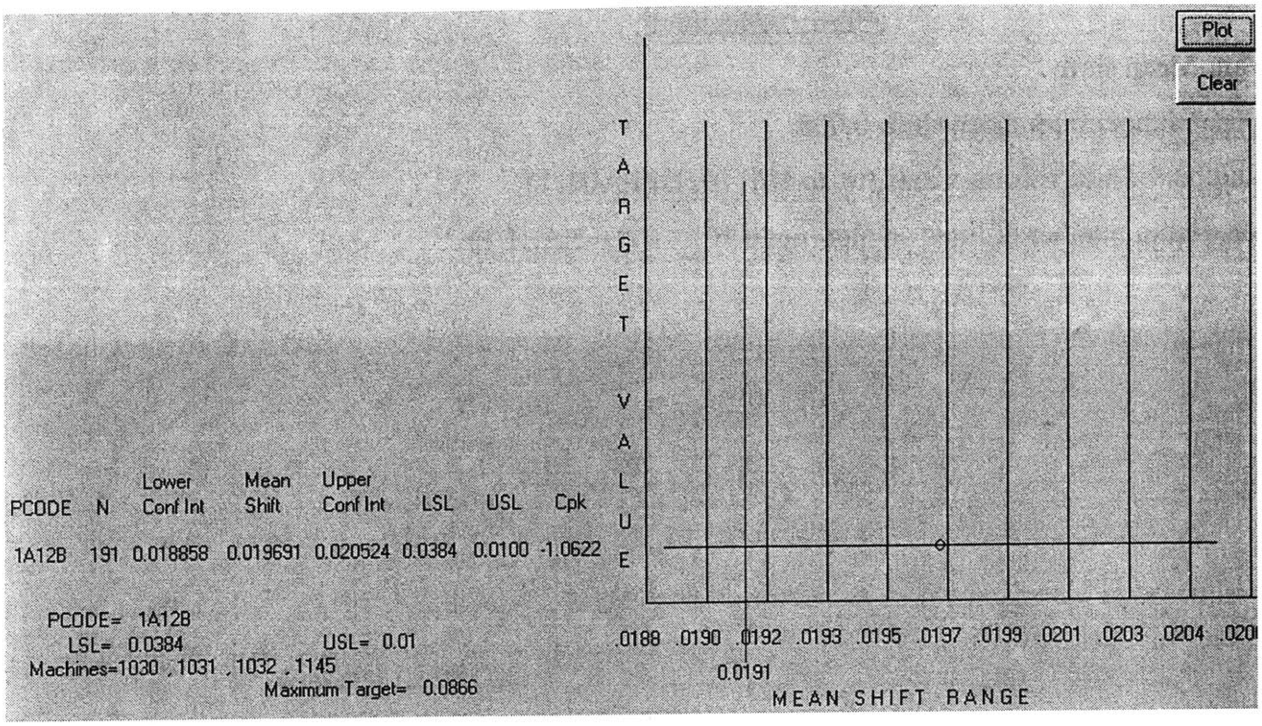


Figure 6.9: Prototype software screen for PCODE 1A12B

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Con
1A12B	0.08660	0.0384	0.0100	1032	191	0.0196911	0.0059	-1.062	0.01885846	0.0205237
1A12B	0.08660	0.0384	0.0100	1145	9	0.1316667	0.2932	-0.138	-0.0599091	0.3232423

Figure 6.10: Data for Figure 6.9

Figure 6.9 shows data exclusion for the PCODE 1A12B. The second line of data in Figure 6.10 is excluded from the plot because it's upper confidence interval is greater than 0.03. This means it is an outlier run.

Form 1 Data Input

Plot: Mean shift

Desired maximum mean shift: 0.008

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

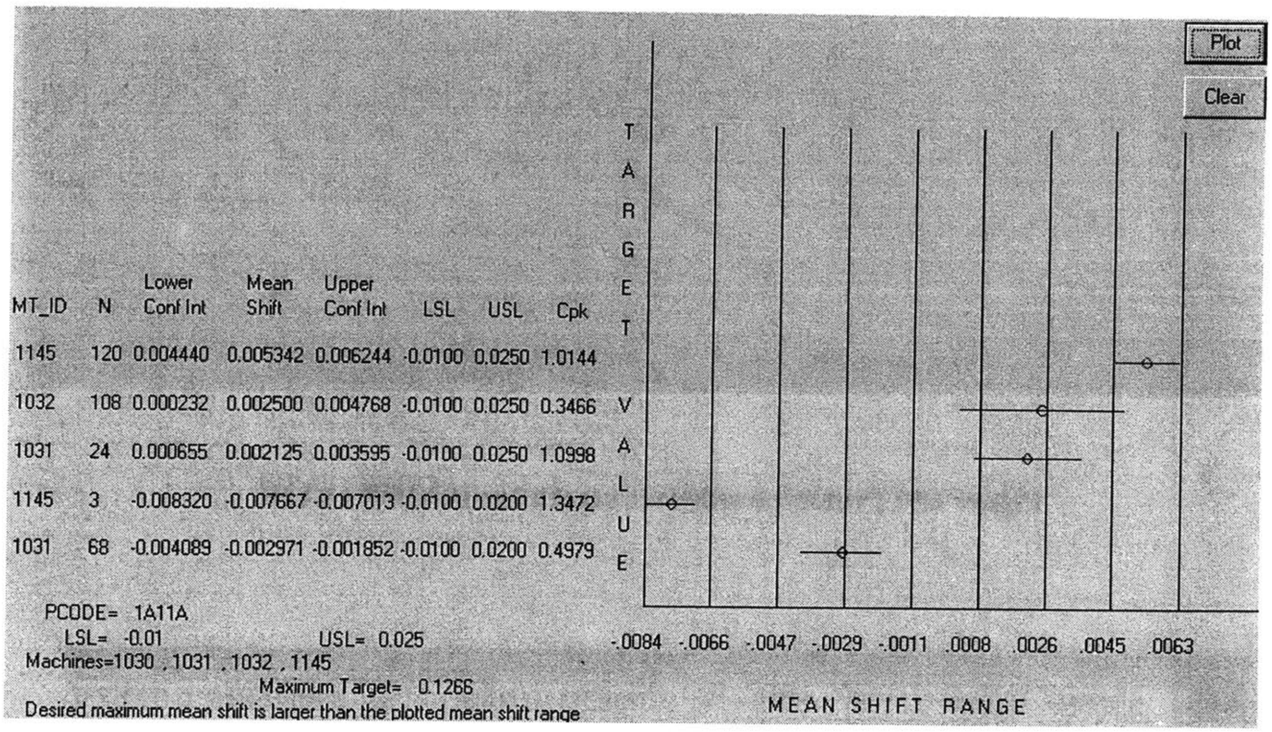


Figure 6.11: Prototype software screen for range of USLs

The sample of data in Figure 6.11 shows how three machines for producing PCODE 1A11A can be compared. It also shows that the user can input a desired number of intervals on the x-axis. Here the user inputted eight; therefore, there are only 8 rather than 10 intervals plotted. Figure 6.11 shows that when the user inputs a USL of 0.025, all USLs between 0.025 and zero are used. Finally, this figure shows that the user can simply input a minimum target and not input a maximum target, if a single target value is desired.

<u>Form 1 Data Input</u>

Plot: Mean shift

Desired maximum mean shift: 0.005 (DEFAULT)

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

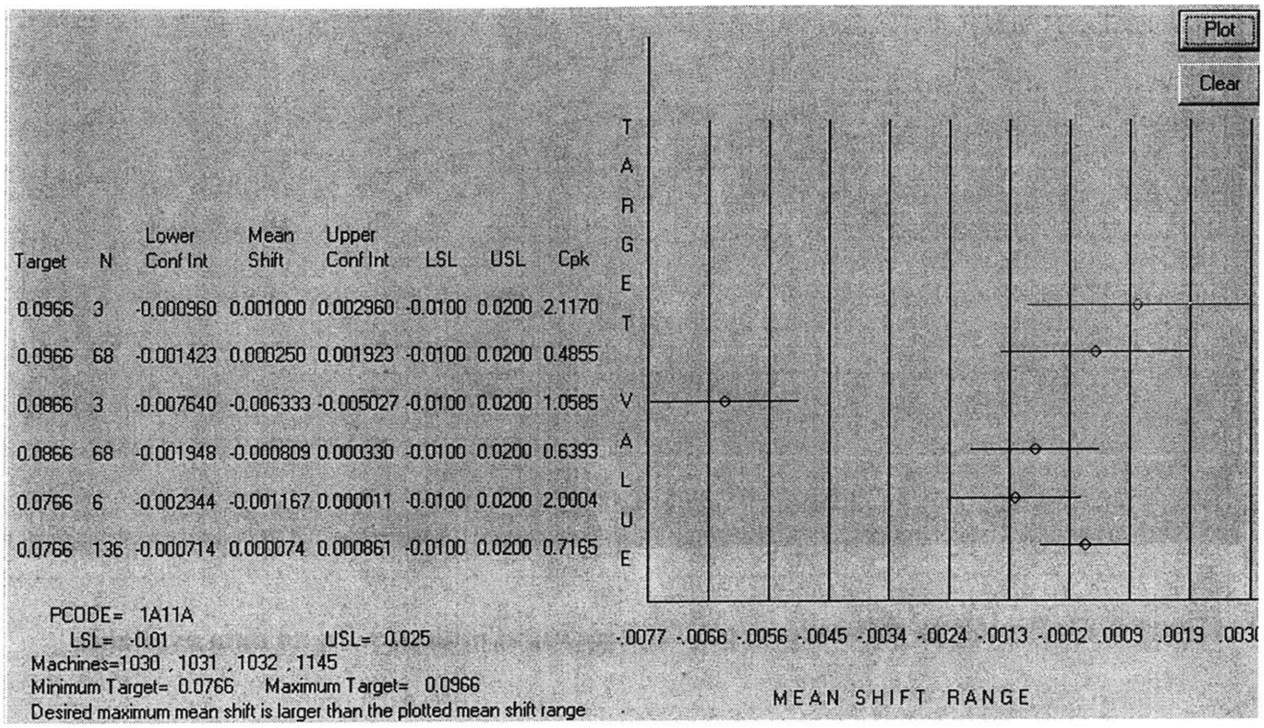


Figure 6.12: Prototype software screen for range of target values

Figure 6.12 shows that a range of target values can be chosen by selecting a maximum and minimum. This is useful if the designer has not yet chosen a dimension.

<u>Form 1 Data Input Form</u>

Plot: Mean shift

Desired maximum mean shift: 0.005 (DEFAULT)

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

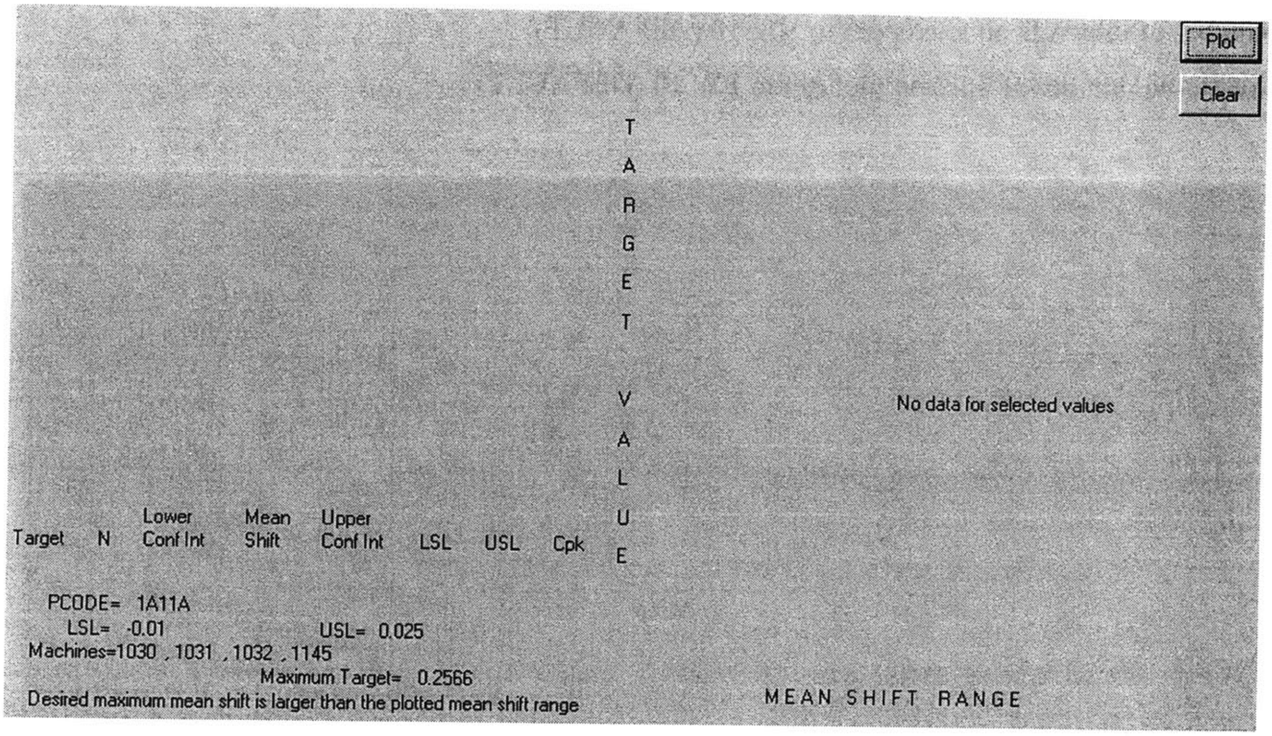


Figure 6.13: Prototype software screen for user input options with no data available

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev	Cpk	Lower Conf	Upper Conf
1A11A	0.25660	-0.0100	0.0250	1032	1	0.026				

Figure 6.14: Data for Figure 6.13

Figure 6.13 shows the plot display which results when the user inputs parameters for which no data is available. Figure 6.14 shows there was one line of data for the user selection of options for each parameter. This data was excluded, however, because it only has one sample.

6.4 Standard Deviation

The following figures show how the prototype software can be used to plot the standard deviation for particular parameter values. There are five examples of using the prototype software to plot standard deviation. The first example shows data exclusion for a standard deviation value that is an outlier. The second example shows how the software plots the user-input desired maximum standard deviation. The third example shows data exclusion for a standard deviation upper confidence interval value that is an outlier. The fourth example shows what the software does when no machines are selected. The fifth example shows how the user can plot standard deviation for a range of target values.

<u>Form 1 Data Input</u>

Plot: Standard deviation

Desired maximum standard deviation: 0.003

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

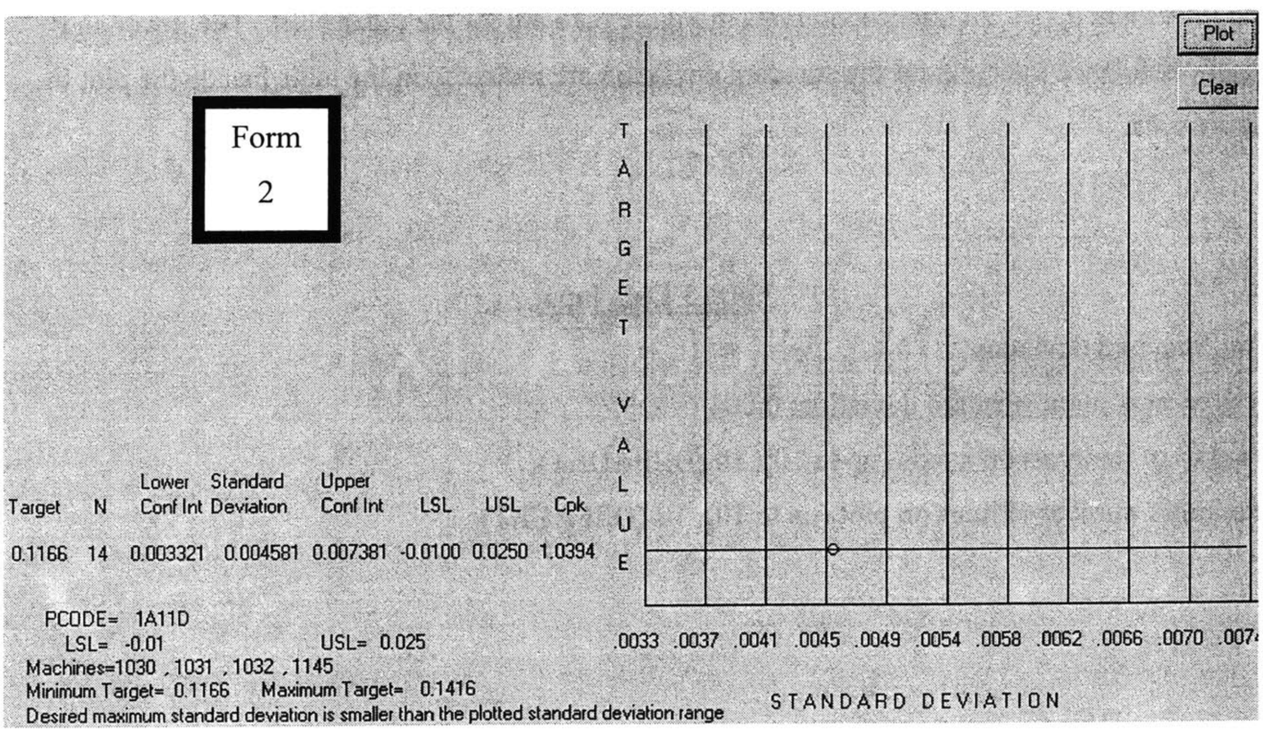


Figure 6.15: Prototype software screen for data exclusion for confidence intervals

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Conf
1A11D	0.11660	-0.0100	0.0250	1032	319	0.00183386	0.049	0.0806	-0.003537	0.00720471
1A11D	0.11660	-0.0100	0.0250	1145	14	0.00428571	0.005	1.0394	0.00188584	0.0066855
1A11D	0.14160	-0.0100	0.0250	1032	93	1.75286022	16.78	-0.034	-1.6584387	5.1641591
1A11D	0.14160	-0.0100	0.0250	1145	3	0.01166667	0.014	0.3079	-0.0046667	0.02

Figure 6.16: Data for Figure 6.15

Figure 6.15 shows how the user can plot standard deviation for a range of target values. It also shows that if the desired maximum standard deviation value is less than the range of the plot, it is indicated to the left of the plot.

Figure 6.16 shows there are four lines of data; however, only one line is plotted. Line 1 is excluded because the standard deviation is greater than 0.015. Line 3 is excluded because the upper confidence interval is greater than 0.0099 and the lower confidence interval is less than (-0.0099). Line 4 is excluded because the upper confidence interval is greater than 0.0099.

The lower and upper confidence intervals in Figure 6.16 are for the mean shift. The lower and upper confidence intervals for the standard deviation are indicated in the table beside the plot in Figure 6.15.

<u>Form 1 Data Input</u>

Plot: Standard deviation

Desired maximum standard deviation: 0.004

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

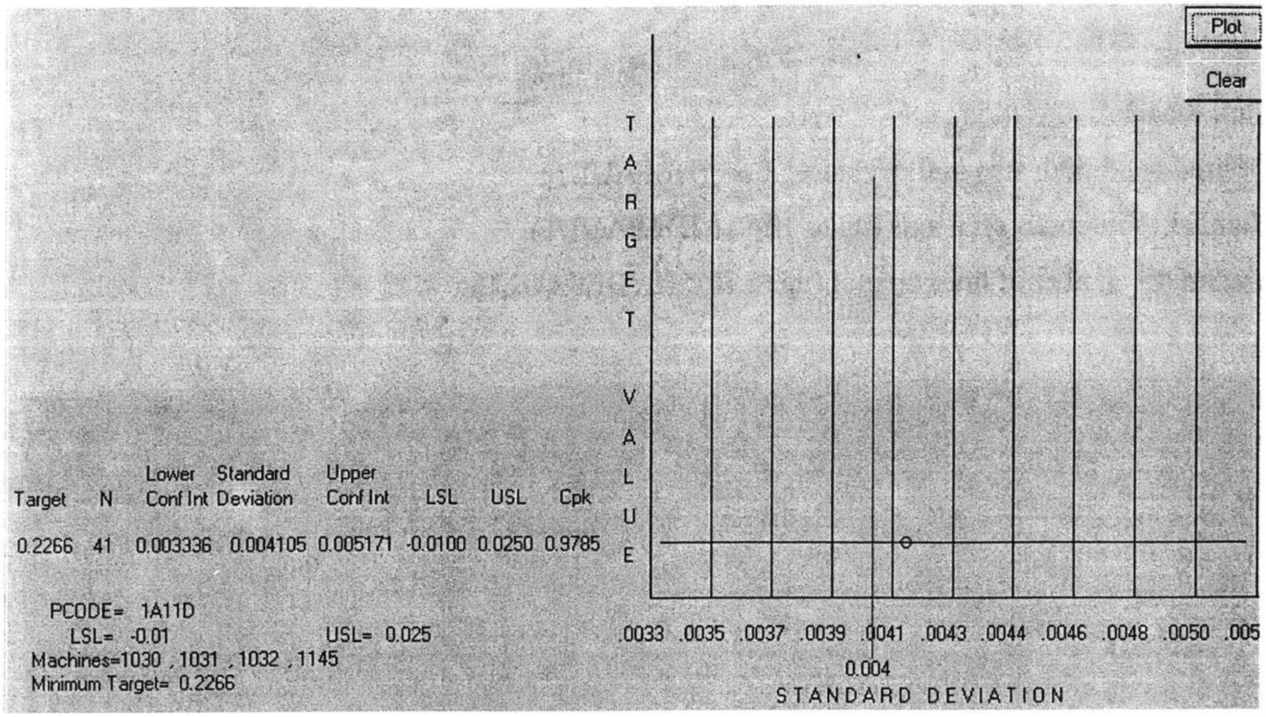


Figure 6.17: Prototype software screen for data exclusion for C_{pk}

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Con
1A11D	0.22660	-0.0100	0.0250	1032	41	0.0020488	0.0041	0.97848	0.00079237	0.00330519
1A11D	0.22660	-0.0100	0.0250	1145	3	0.005	3E-09	1643820	0.005	0.00500000

Figure 6.18: Data for Figure 6.17

Figure 6.17 shows how standard deviation PCD can be used to compare how various machines produce PCODE 1A11D. Figure 6.18 also shows that the user input value for the desired maximum standard deviation is plotted.

Figure 6.18 shows two lines of data, but only one line is plotted. Line 2 is not plotted because C_{pk}>5.0. The lower and upper confidence intervals in Figure 6.18 are for the mean shift. The lower and upper confidence intervals for the standard deviation are indicated in the table beside the plot in Figure 6.17.

<u>Form 1 Data Input</u>

Plot: Standard deviation

Desired maximum standard deviation: 0.01 (DEFAULT)

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

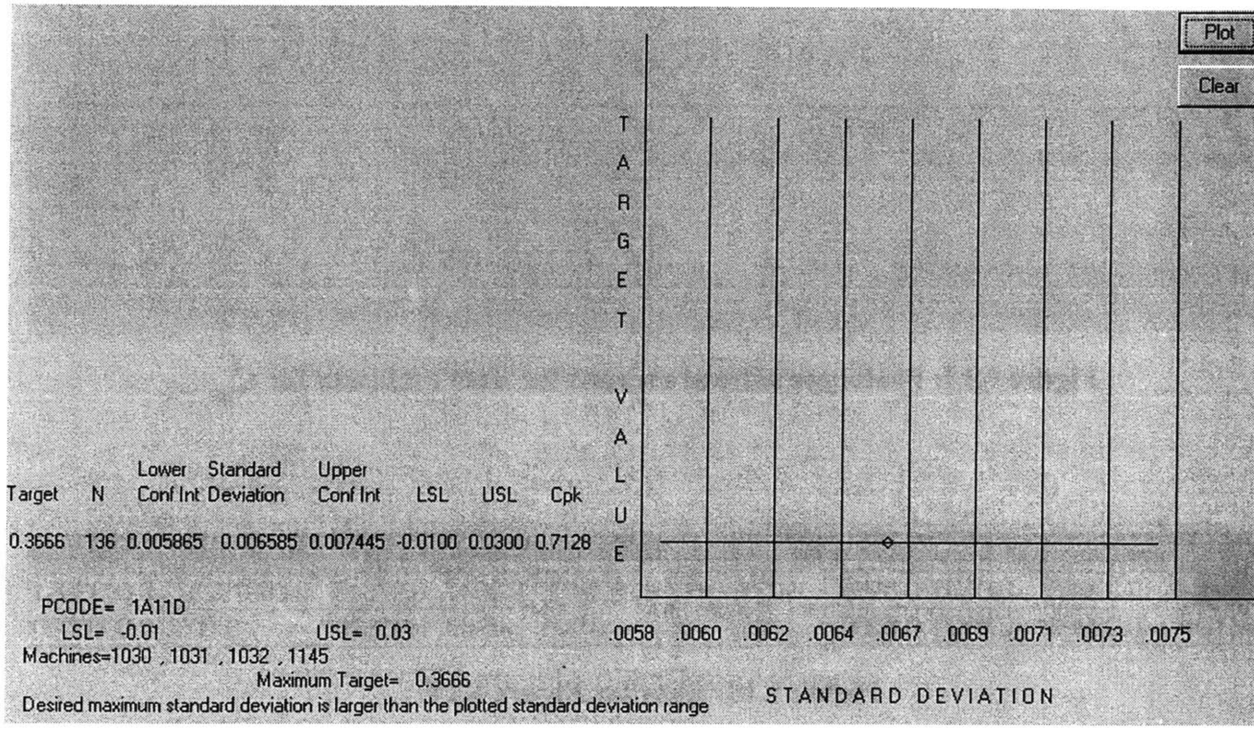


Figure 6.19: Prototype software screen for data exclusion for upper confidence interval

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Con
1A11D	0.36660	-0.0100	0.0300	1031	136	0.00408088	0.0066	0.71276	0.00297413	0.00518763
1A11D	0.36660	-0.0100	0.0300	1145	6	0.02216667	0.001	2.65575	0.02137995	0.02295338

Figure 6.20: Data for Figure 6.19

Figure 6.19 shows how the standard deviation can be plotted for several machines for producing 1A11D parts. It also shows that the user can input the number of intervals to plot on the x-axis.

There are eight intervals in this example. Finally, this figure indicates that the desired maximum standard deviation is greater than the range plotted.

In Figure 6.20, there are two lines of data; however, only line 1 is plotted because the maximum allowable upper confidence interval is exceeded by line 2. The lower and upper confidence intervals in Figure 6.20 are for the mean shift. The lower and upper confidence intervals for the standard deviation are indicated in the table beside the plot in 6.19.

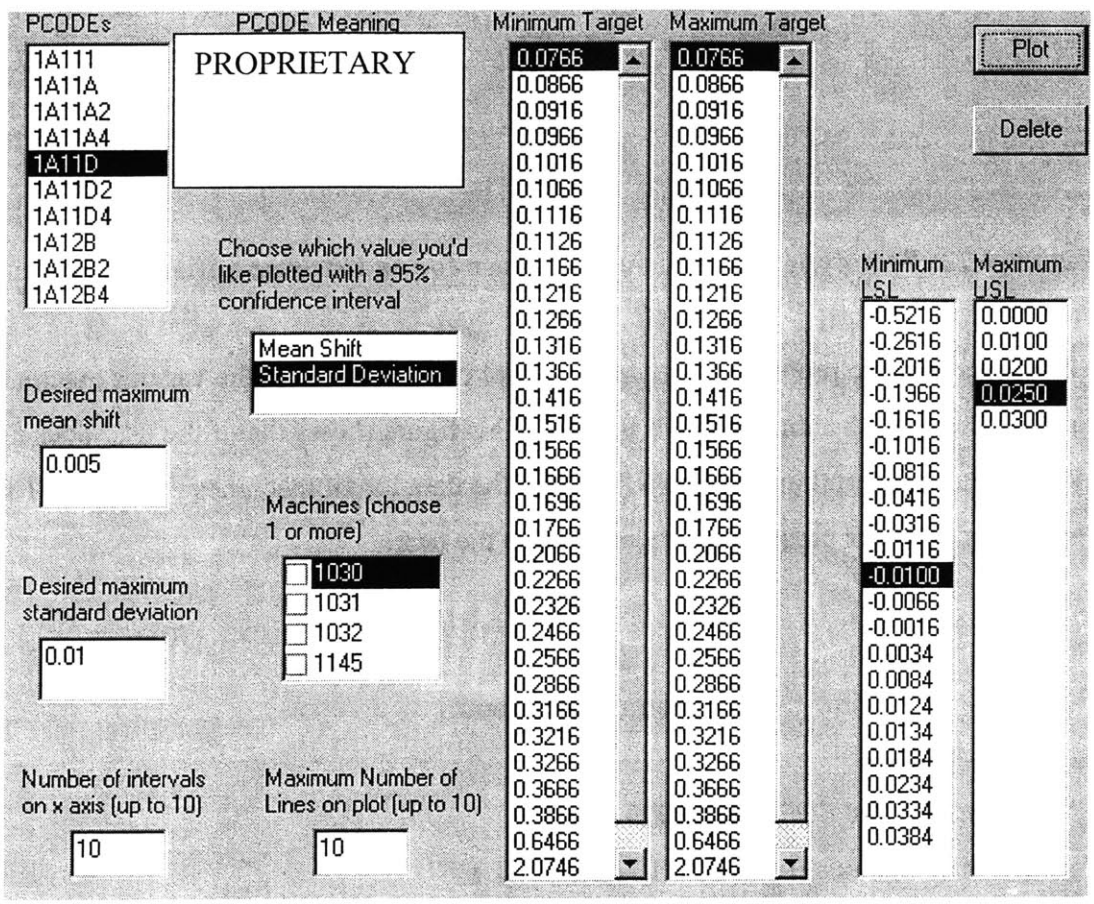
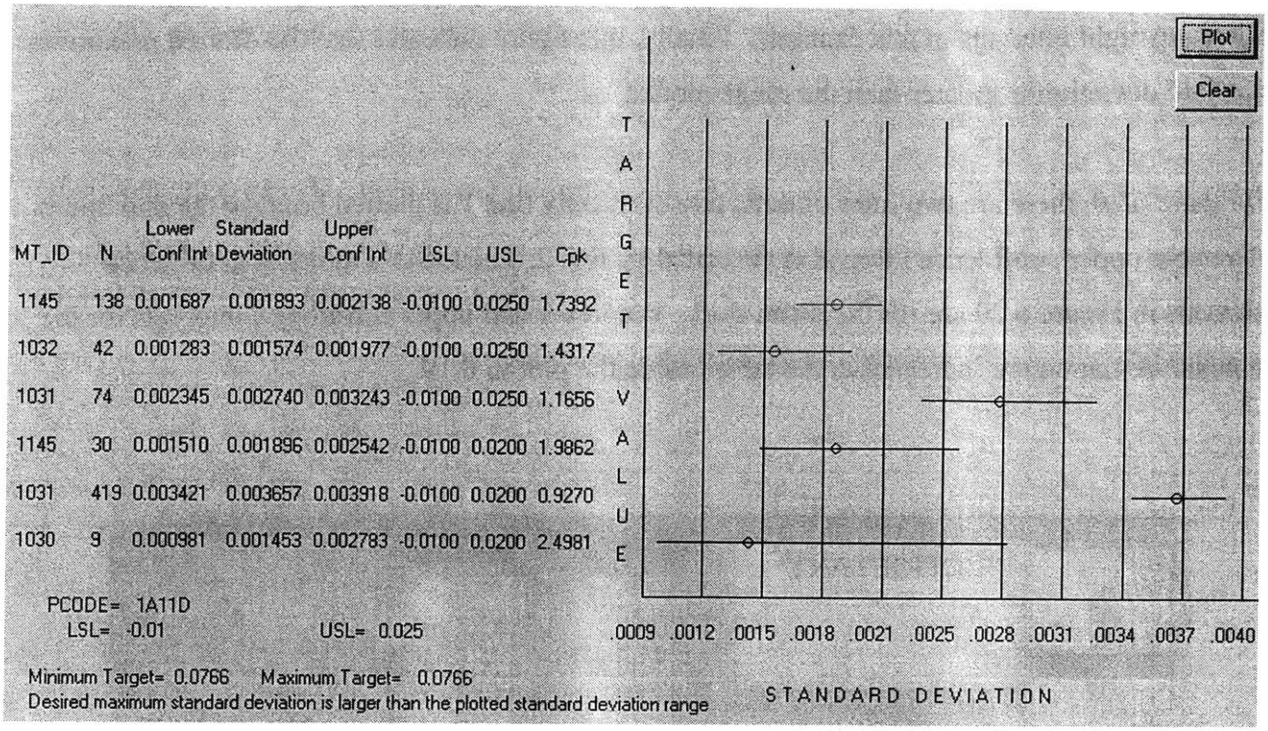


Figure 6.21: Prototype software user input form screen for no machine options chosen



MT_ID	N	Lower Conf Int	Standard Deviation	Upper Conf Int	LSL	USL	Cpk
1145	138	0.001687	0.001893	0.002138	-0.0100	0.0250	1.7392
1032	42	0.001283	0.001574	0.001977	-0.0100	0.0250	1.4317
1031	74	0.002345	0.002740	0.003243	-0.0100	0.0250	1.1656
1145	30	0.001510	0.001896	0.002542	-0.0100	0.0200	1.9862
1031	419	0.003421	0.003657	0.003918	-0.0100	0.0200	0.9270
1030	9	0.000981	0.001453	0.002783	-0.0100	0.0200	2.4981

Figure 6.22: Prototype software output screen for no machine options chosen

Figure 6.22 shows how the prototype software could be used to compare the various machines to produce a 1A11D part with a dimension of 0.0766. This figure shows that if the user does not chose any of the machine options (Figure 6.21), then the data for all machines is plotted. This is how the prototype software presents aggregate data to the user.

<u>Form 1 Data Input</u>

Plot: Standard deviation

Desired maximum standard deviation: 0.003

Number of intervals on x axis (up to 10): 10 (DEFAULT)

Maximum number of lines on plot (up to 10): 10 (DEFAULT)

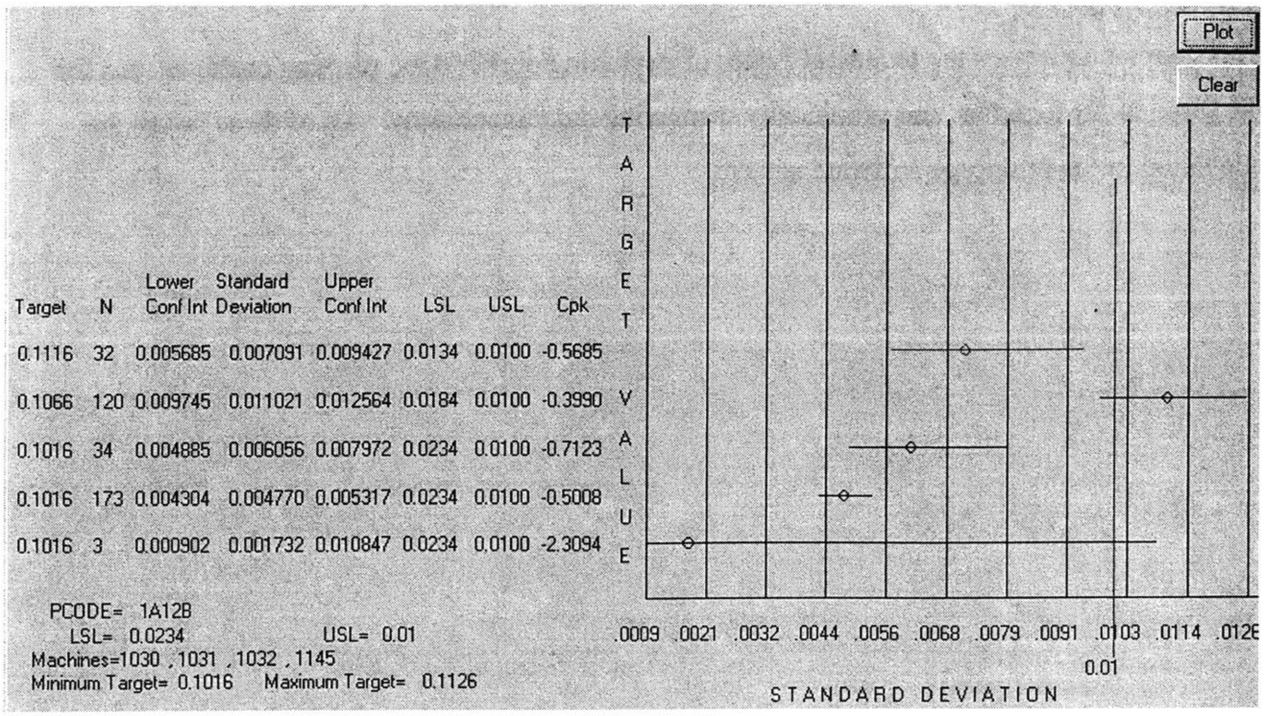


Figure 6.23: Prototype Software screen for range of targets

PCODE	TARGET	LSL	USL	MT_ID	N	MeanShift	StDev(	Cpk	Lower Conf	Upper Conf
1A12B	0.10160	0.0234	0.0100	1031	3	0.022	0.002	-2.3094	0.02004	0.02396000
1A12B	0.10160	0.0234	0.0100	1032	173	0.0171676	0.005	-0.5008	0.01645677	0.01787849
1A12B	0.10160	0.0234	0.0100	1145	34	0.0229412	0.006	-0.7123	0.02090544	0.02497691
1A12B	0.10660	0.0184	0.0100	1032	120	0.0231917	0.011	-0.399	0.02121972	0.02516361
1A12B	0.10660	0.0184	0.0100	1145	9	0.0292222	0.010	-0.6153	0.02241863	0.03602582
1A12B	0.11160	0.0134	0.0100	1032	32	0.0220938	0.007	-0.5685	0.01963687	0.02455063
1A12B	0.11260	0.0124	0.0100	1032	32	0.0191875	0.016	-0.1885	0.01355890	0.0248161

Figure 6.24: Data for Figure 6.23

Figure 6.24 shows that the data shown in Figure 6.23 is for multiple targets, multiple LSLs, and multiple machines. It also shows that when the user chooses a LSL of 0.0234, all LSLs between 0.0234 and zero are plotted.

6.5 Conclusion

This chapter addresses the technical issues of excluding outlier data, plotting multiple runs for the same index together, and graphically displaying data uncertainty. All of these issues are alleviated in the prototype software system.

7 Hierarchy of PCDB

7.1 *Introduction*

This chapter addresses two of the technical barriers to design usage of PCD. First, methods for developing a consistent PCDB classification scheme are discussed. Second, quantitative methods for determining surrogate data for unpopulated indexes are presented.

This chapter discusses the process capability database hierarchy of the large aerospace company studied. Section 7.2 describes the current hierarchy. Four problems with this hierarchy were noted. First, traversing the hierarchy is complicated because the user needs to choose options for many parameters (Section 7.3). Second, the classification scheme is inconsistent (Section 7.4). Third, it is difficult to determine surrogate data for the vast number of unpopulated index options (Section 7.5). Finally, the perpetually changing of the hierarchy is a problem (Section 7.6).

7.2 *Infeasible indexes*

In the large aerospace company's PCDB a user would need to input a feature, a material, and a process in order to obtain specific data. A user would progress through a set of choices for each of these parameters based on each previous option that he/she had selected. Each option selected defined a further digit in the PCODE. For example, if the user selected Aluminum alloy as the material, which is PCODE 1.2.1, he/she then had to select the specific type of Aluminum, which represented the fourth digit of the PCODE. The user had to go through this parameter selection for each of the material, the feature, and the process.

7.2.1 Problem

The problem with the hierarchy used by the large aerospace company was that it did not eliminate infeasible options. Figures 7.1 and 7.2 show what the screen progressions look like for the selection of the feature of rib/stiffener length. Each separate box represents a new input screen. The various input screens are used in progression for PCD selection.

In Figure 7.1, only the options that are highlighted in light gray in the second box are applicable to "Feature". The other options are for "Process" and "Material". In the large aerospace company's system, all of these options were listed even though only the feature options were feasible. In Figure 7.2, only the options that are highlighted in light gray in the third box are feasible choices for the "Rib/stiffener" feature. All the other choices are applicable only for other feature choices such as "flange" or "hole".

Figures 7.1 and 7.2 show that as the user chose more detailed information about the data desired, the list of options on progressive screens became more expansive. The large aerospace company has 52 million possible indexes only considering defined positions in the eight-digit ndex and only considering machining processes. Of these potential indexes, there are only approximately 50,000 feasible index combinations (0.1%).

1. Select: Material **Feature** Process

First digit = 3

2. Select:		
Rib/stiffener	Flange	Hole
Countersink	Counterbore	…
…	Radius	Slot
Detail part	…	…
Chamfer/taper	…	…
Complex contour	Simple contour	Fastener
…	Web/floor	…
Threads	…	…
Boss	…	Inside Diameter
Outside diameter	Adhesives	Alloys
…	Chemicals	…
Molding Materials	…	…
…	…	Environmental Ctrl
Aluminum Titanium	Magnesium	…
…	…	Clean
…	…	…
Hole Preparation	Fastener Installation	Forming
…	Milling	Molding
…	…	Punch/Blank
…	Vendor	Paint
….	…	…
Saw	…	…
…	Chamfer	…
Deburr	Surface Blend/Finish	…
Precision Shape	…	…
…	…	Honing
Turning	Grinding	Boring
…	…	Weld
Threads	…	Office

2^{nd} digit = 1

Selected Item Feasible items based on choice of "Feature"

… represent places where there are other options that cannot be listed for proprietary reasons

Figure 7.1: Current screen progression for data selection for feature and PCODE 3.1

2. Select:	**Rib/stiffener**	Flange	Hole
	Countersink	Counterbore	…
	…	Radius	Slot
	Detail part	Cut out	…
	Chamfer/taper	Step/joggle	…
	Complex contour	Simple contour	Fastener
	…	Web/floor	…
	Threads	…	…
	Boss	…	Inside Diameter
	Outside diameter		

Second digit = 1

⇩

3. Select:	Location	Thickness	**Length**
	Height	Angle	Profile
	Width	Diameter	…
	…	Quantity	…
	Depth	…	Outside diameter
	Inside diameter	…	Size
	…	…	Overall calculated length
	Surface finish	…	Overall calculated width
	…	Weight	Overall calcul Thickness
	…	…	…
	Pressure Check	Offset	…
	…	Hole-to-hole	…
	…	…	…
	…	…	…
	Material defect	…	Fastener head diameter
	Time	Temperature	Fastener shank diameter
	…	…	…
	Air pressure	…	Viscosity
	Failures	…	…
	Paint thickness	Primer thickness	…
	…	…	Environmental Control
	…	…	…
	…	Vendor	…
	% elongation	…	…
	Hardness	…	…
	…	Yield strength	Ultimate strength
	…	…	…
	End Mill	…	…
	…	…	…

Third digit = 3

Selected Item Feasible item based on selection of "Rib/Stiffener"

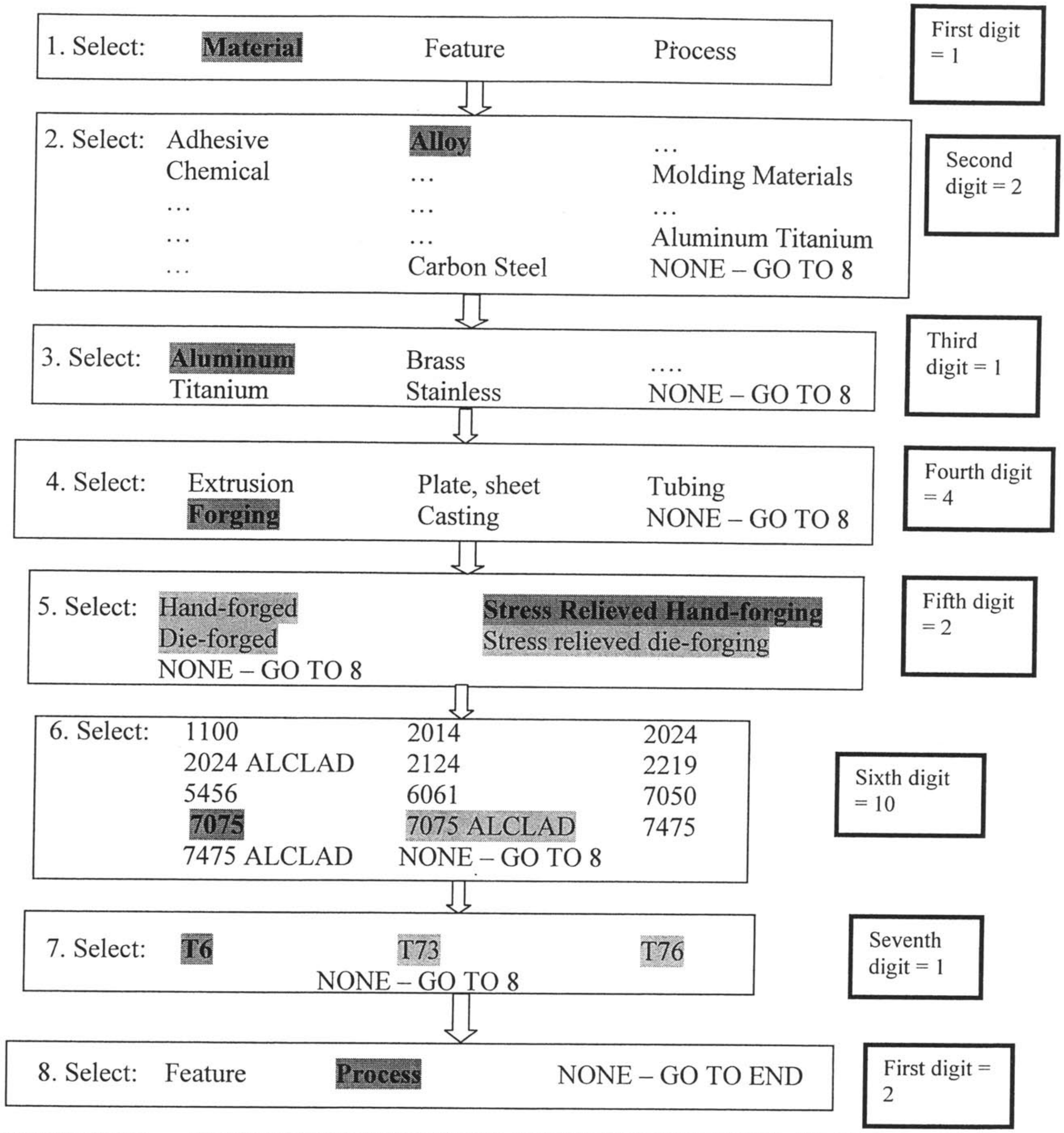
… represent places where there are other options that cannot be listed for proprietary reasons

Figure 7.2: Current screen progression for data selection for feature and PCODE 3.1.3

7.2.2 Proposed solution

As a solution to this problem of infeasible index listings, an alternative set of option screens is proposed. This new scheme simply eliminates all options that are infeasible based on the options that have already been chosen. Some of the options should be eliminated entirely such as "Office", "Vendor", and "Environmental Control" in the second block of Figure 7.1. These options do not have a defined process capability. An example of the proposed set of option screens in shown in Figure 7.3.

Figure 7.3 shows an example of the user choosing a material and more detailed options for the specific type desired. Again each box represents a new screen and they are numbered consecutively in their progression order. The bold text with the dark highlighting represents the parameter that was chosen and the light highlight represents alternative parameters that could have been selected that would have resulted in the same following screen. Whenever the user has inputted the amount of detail he/she desires, he/she can choose the "NONE" to proceed to screen 8 to choose either "process" or "feature" to input next in a similar screen progression. Not inputting all the details for each of the three parameters would result in the need for aggregate data, which is discussed in Section 5.4.

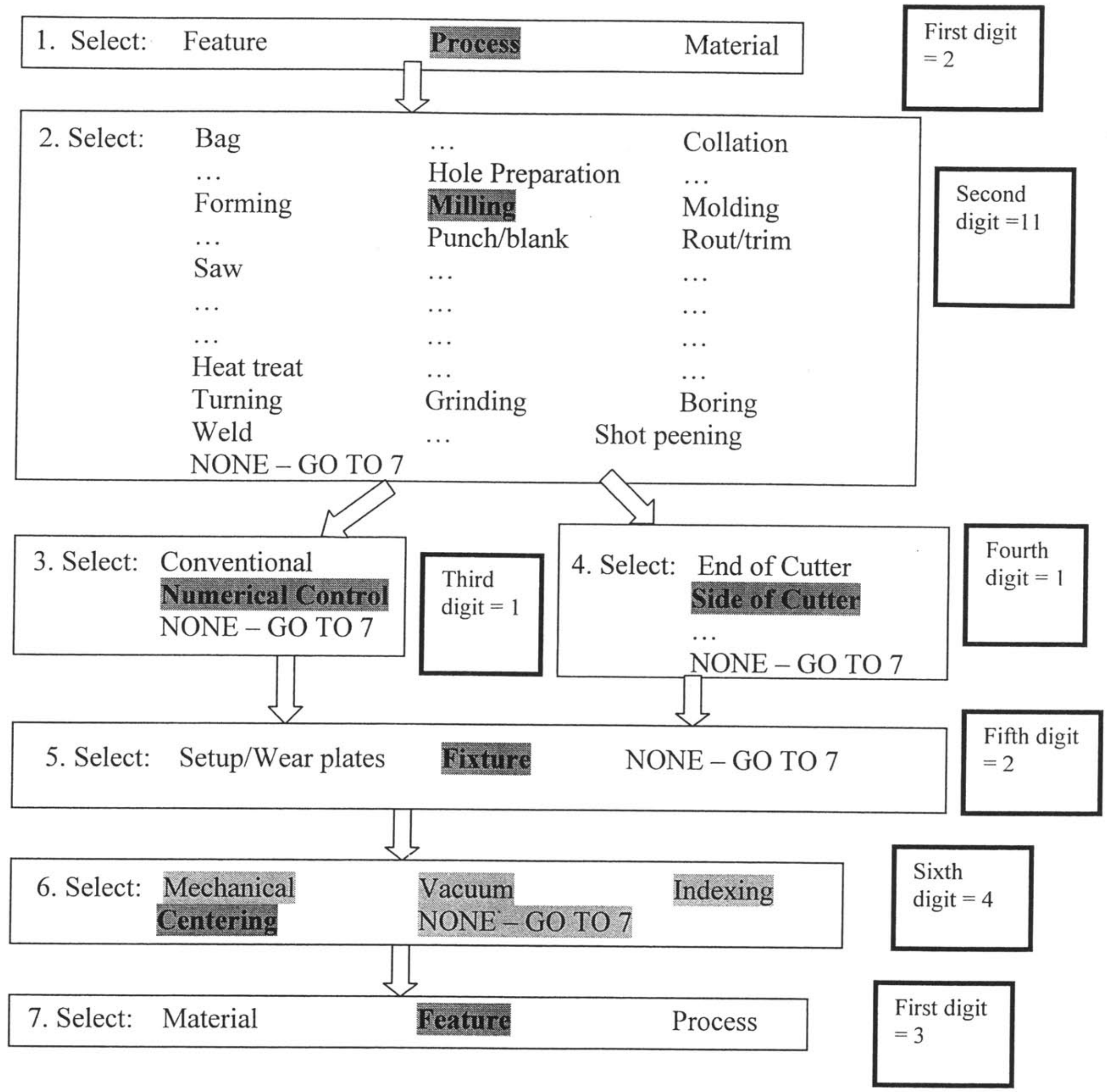


Selected Item   Alternative items that would result in the same following screen based on previous selected options

… represent places where there are other options that cannot be listed for proprietary reasons

Figure 7.3: Proposed screen progression for data selection for material and PCODE 1.2.1.4.2.10.1

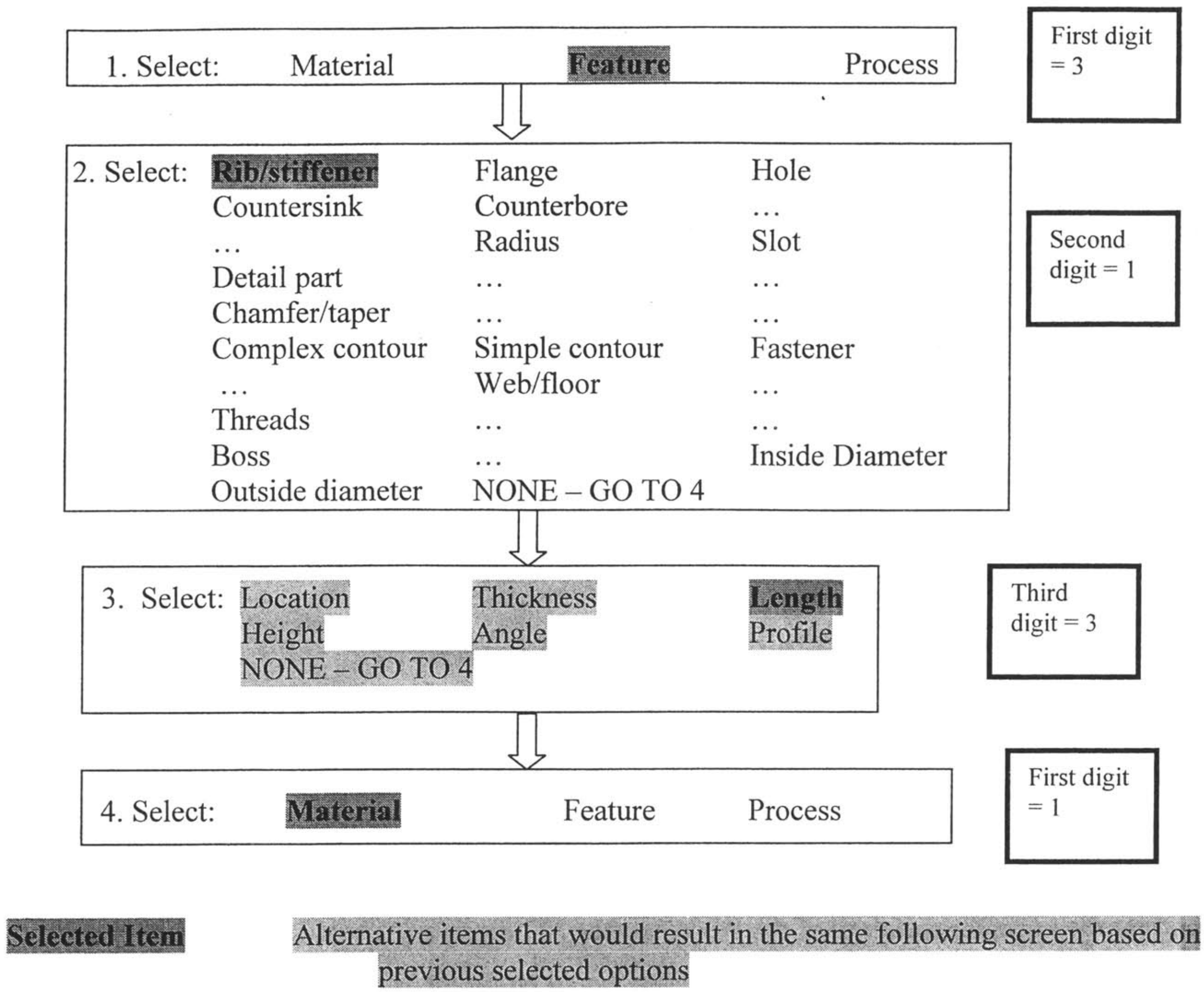
Another example of what this screen progression would look like for a process selection is shown in Figure 7.4 and another for a feature selection in Figure 7.5. Figure 7.5 can be compared to Figures 7.1 and 7.2. In Figure 7.5, after "Feature" is selected there are only 31 options available and after selecting "Rib/stiffener", there are only there. On the other hand, there are 87 options in Figure 7.1 and 56 options in Figure 7.2 for these same selections. In Figure 7.5, the user can only choose feasible indexes whereas in Figures 7.1 and 7.2, the probability of an infeasible index being selected is 43.4% and 96.1% respectively.



Selected Item   Alternative items that would result in the same following screen based on previous selected options

… represent places where there are other options that cannot be listed for proprietary reasons

Figure 7.4: Proposed screen progression for data selection for process and PCODE 2.11.1.2.4.3



… represent places where there are other options that cannot be listed for proprietary reasons

Figure 7.5: Proposed screen progression for data selection with current hierarchy scheme for feature and PCODE 3.1.3

7.3 Inconsistent classification scheme

The inconsistency problem with the hierarchy of the large aerospace company results from having multiple meanings for one number in a particular digit of the PCODE. For each PCODE digit, each value that can be used should have a unique meaning.

7.3.1 Problem

The hierarchy at the large aerospace company is set up like a tree with each further branch from the trunk being a further digit in the PCODE. The problem is that parallel branches are not similar. Figure 7.6 shows what the various digits of the PCODE mean.

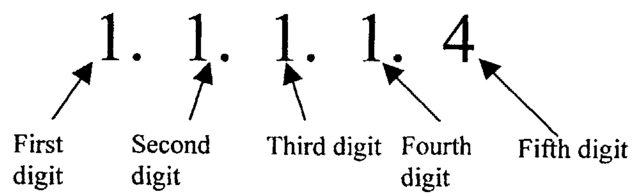


Figure 7.6: PCODE digit explanation

Figure 7.7 shows what part of the current system looks like where the tree is "Adhesives" and this flows down to branches such as "Film" and "Foam". The numbers listed in each box represent the value for that digit of the PCODE. The "1" digit for Adhesive would be for the first digit in the PCODE. The "1" digit for Film would be for the second digit in the PCODE.

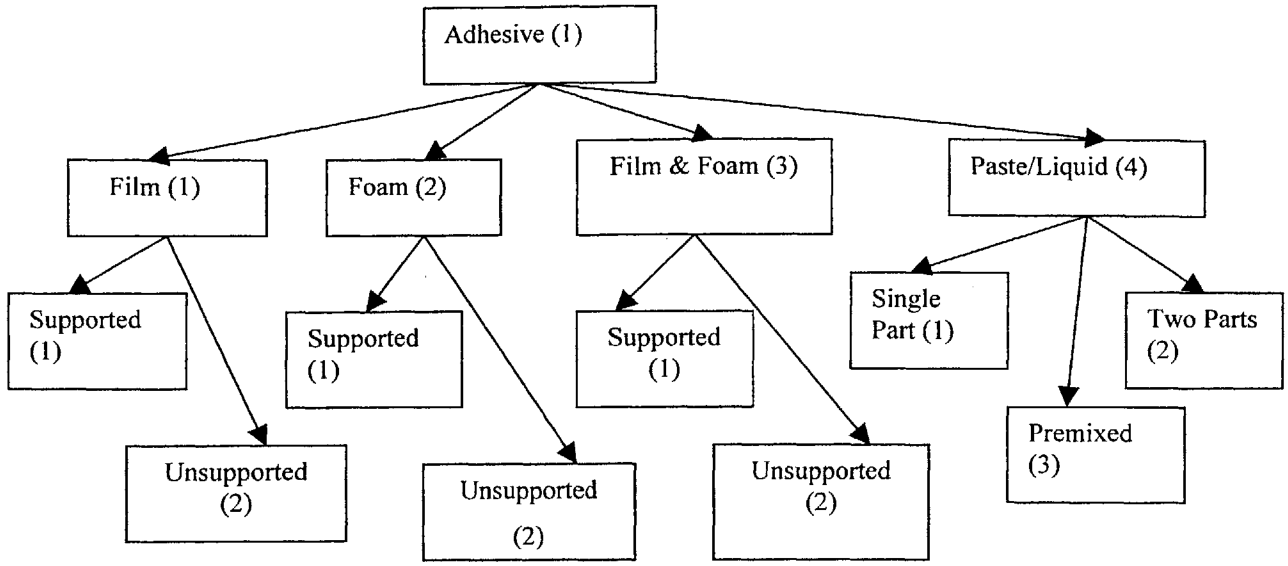


Figure 7.7: Different meaning of the same number in the same digit location

The problem arises because for foam, film, and film & foam, the third digit of one represents "supported" and the third digit of two represents "unsupported". Nonetheless, for paste/liquid, the third digit of one represents "Single part" and the third digit of two represents "Two parts". This is a problem when looking for surrogate data. It is assumed that indexes with the same number in the same digit location are similar; however, here the third digit representation of the numbers one and two have dramatically different meanings. It is possible that 1.1.1 is similar to 1.2.1 and 1.3.1, but it is probably dissimilar from 1.4.1.

Also, if aggregate data is needed for adhesives and the choice of foam, film, film and foam, or paste/liquid has not been made yet, the data for all of these will be presented together. The data for paste/liquid should not be combined with the data for film, foam, and film and foam to obtain aggregate data for adhesives.

7.3.2 Proposed solution

To remedy this problem it is suggested that each number within each digit location have its own unique meaning. Figure 7.8 shows what this would look like for the same PCODEs as shown in

Figure 7.7. The PCODEs 1.4.1, 1.4.2, and 1.4.3 would be replaced by 1.4.3, 1.4.4 , and 1.4.5 respectively.

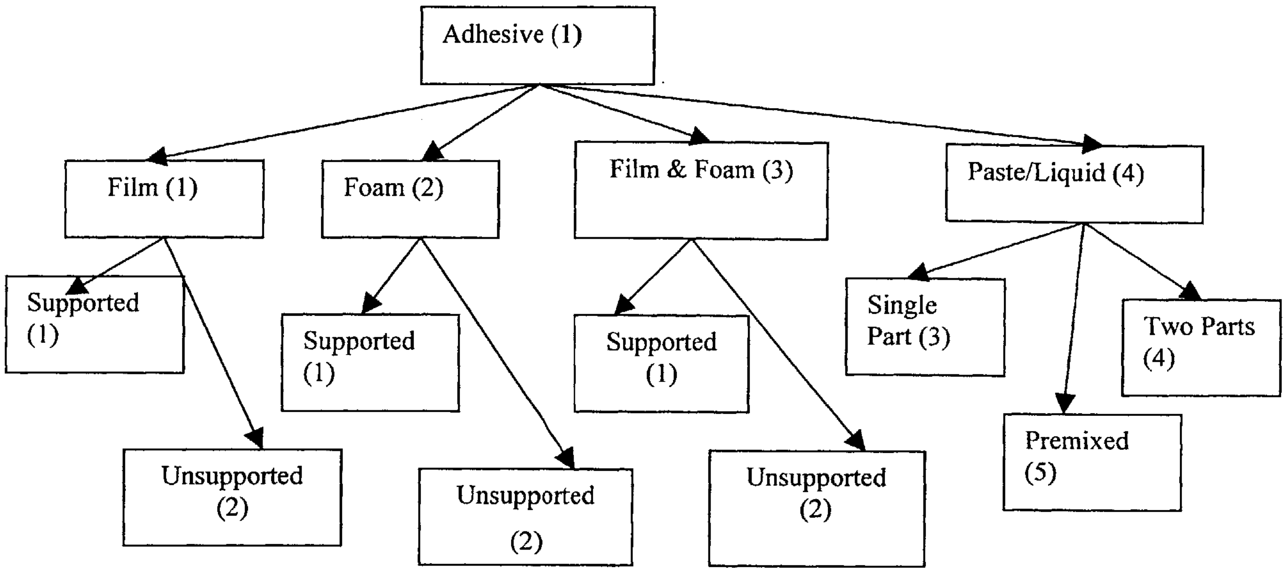


Figure 7.8: Proposed unique numerical scheme for each digit location

7.4 Surrogate data for unpopulated indexes

Unpopulated databases were identified as a barrier to process capability data usage by design. As the indexing schemes of the databases are becoming more detailed, more gaps of unpopulated data results. The large number of indexes makes it highly unlikely that an exact match will be found. If there are 10 process, 10 feature, and 10 material options in the index, then there are 1,000 possible indexes. The potential size grows exponentially with more detail. With the increased size, there is a greater possibility that a designer may request process capability for a index that has no data. For the approximately 50,000 feasible indexes for the large aerospace company, only about 487 are populated with at least 10 runs of data.

7.4.1 Problem

If a designer wants data for a particular material, process, and feature and this data is missing from the database (unpopulated), alternative (surrogate) data must be chosen. The surrogate data should be taken from the index that is most similar to the unpopulated index. The most similar data might be for the same process, feature, or material or any combination of one or two of these three parameters. For example, if a designer would like tolerance information on drilling small holes in Aluminum and this specific data is missing from the PCDB, should he/she utilize the data for drilling small holes in Titanium, for drilling large holes in Aluminum, or for punching small holes in Aluminum?

There are three alternative methods to determine which surrogate index should be used when the desired index is unpopulated. The first is a predetermined matrix that finds alternative data that shares a majority of the same indexing information. The second is expert knowledge of manufacturing engineers. The third is a quantitative analysis of the data to find the most mathematically similar data. The feasibility, benefits, and shortcomings of the three alternatives need to be evaluated before the best decision is selected. This is future work; however, some ideas have been formulated.

A chart could be developed to help determine surrogate data for unpopulated PCODEs. This chart could be provided directly through the PCDB so that designers can utilize it, rather than VR coordinators, to determine surrogate data. The format of this chart is shown in Table 7.1.

Possible surrogates	P.S. 1	P.S. 2	P.S. 3	P.S. 4	P.S. 5	P.S. 6
Comparison to index without data	0		+	0	-	+

+ indicates the surrogate index is similar to the unpopulated index, but always slightly better. This provides the upper bound of the process capability for the unpopulated index

- indicates the surrogate index is similar to the unpopulated index, but always slightly worse. This provides the lower bound of the process capability for the unpopulated index

0 indicates there is no similarity between the surrogate and unpopulated indexes.

BLANK indicates the surrogate index is also unpopulated.

Table 7.1: Substitute data chart

7.4.2 Proposed solution - quantitative method

One quantitative method has been developed which can be used to determine appropriate surrogate data. This method involves a calculation of a Z value. This is the same Z value presented in Section 5.2.3.

$$z = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \qquad (1)$$

Null Hypothesis: $H_0: \mu_1 - \mu_2 = \Delta_0$

This hypothesis is correct if $-z_{\alpha/2} \leq z \leq z_{\alpha/2}$

Where $\bar{x}$ is the average value for sample 1, $\bar{y}$ is the average value for sample 2, m is the number of runs in sample 1, n is the number of runs in sample 2, Δ_0 is the null value, σ_1 is the standard deviation of sample 1, σ_2 is the standard deviation of sample 2, and α is 0.05 for a 95% confidence interval.

The hypothesis is that the two samples are similar enough to be surrogates for each other, which means Δ_0 is zero. This results in the following formula for the Z value:

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \qquad (2)$$

For a 95% confidence interval, if Z > 1.96 or Z < (-1.96) then sample 2 is a feasible surrogate for sample 1.

The problem with using the Z value quantitative method is that there is no data to use for the unpopulated index. Instead one must look at PCODEs that are similar to the unpopulated one. For these similar PCODEs, the Z value can be used to look for data that is similar enough that it could be used as surrogate data. Then, it can be determined which digits of the PCODE are the most significant for determining appropriate surrogate data.

An example of utilizing this quantitative method is detailed in Section 7.4.3. Figure 7.9 shows a flow diagram of the quantitative process to determine surrogate data. A program could be designed to perform the steps in this flow chart to determine optimal surrogate data for unpopulated indexes.

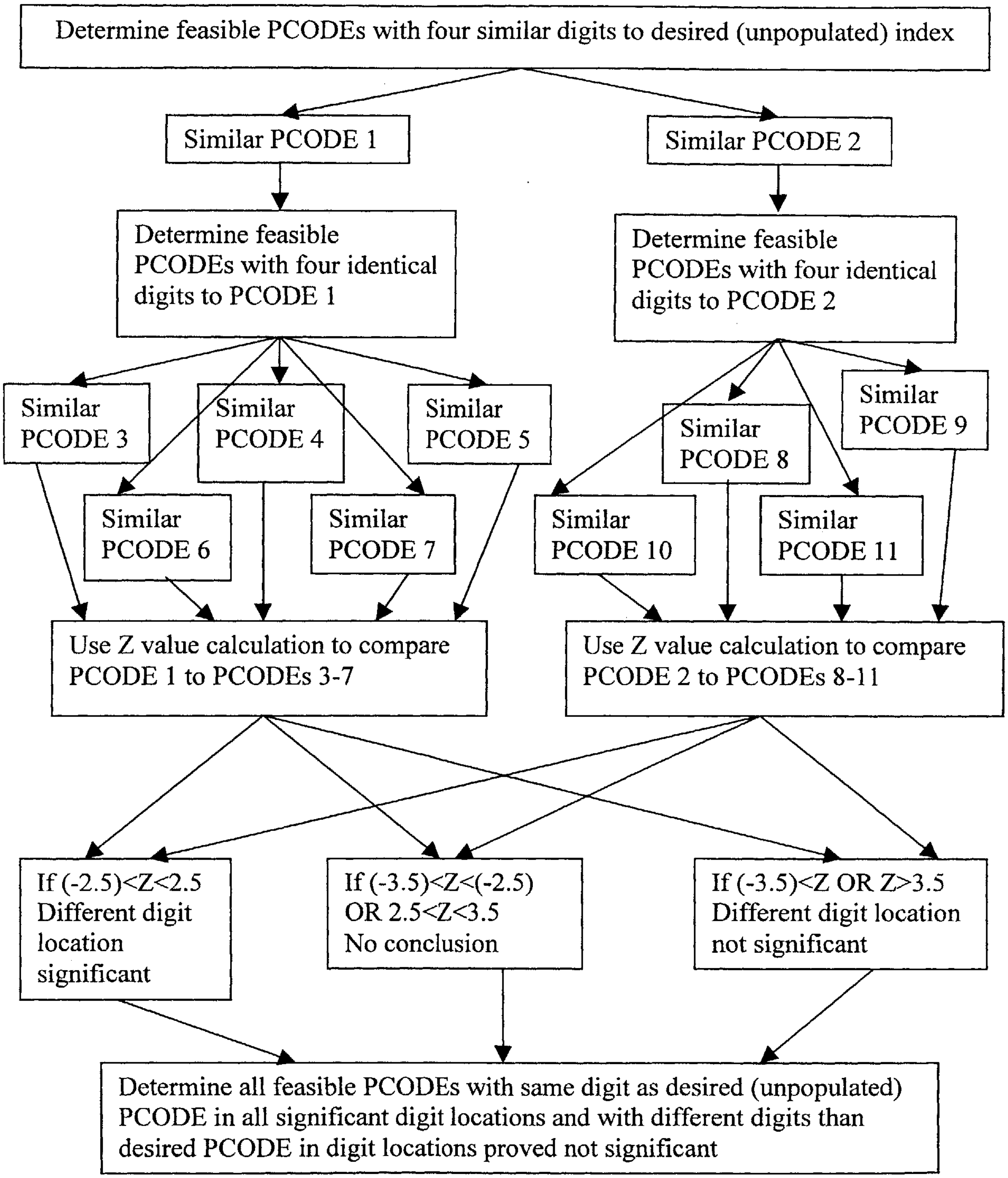


Figure 7.9: Flow diagram for using quantitative method to determine surrogate data

7.4.3 Example of proposed solution

If a designer wants information on the PCODE 1.1.1.1.4 (Film adhesive supported cured-to-cured composite) and this data is not populated in the database, what surrogate data should be used? In this example, alternative PCODEs are examined as potential surrogates. The other parameters (specification limits, target value, and machine) are constant such that the surrogate index would have the same values for these parameters as the unpopulated index.

For this example where data is needed for unpopulated PCODE 1.1.1.1.4, the quantitative program would first look at the PCODE 1.1.1.1.1 which has the first four digits the same as for the desired PCODE (1.1.1.1.4) and would determine which PCODE is the most similar to it. The comparison is in Table 7.2.

PCODE						Average	Std dev	Z (30 samples)	Most Alike 1.1.1.1.1	Similar digits
1	1	1	1	1		1	0.04			
1	1	1	2	1		1.02	0.03	3.286335345	3	1, 2, 3, 5
1	1	1	1	2		1.03	0.015	-3.846364589	4	1, 2, 3, 4
1	1	3	1	1		0.97	0.02	-2.449489743	2	1, 2, 4, 5
1	1	2	1	1		0.99	0.035	1.03050808	1	1, 2, 4, 5
1	1	1	1	3		1.038	0.032	-4.063144883	5	1, 2, 3, 4

1.1.1.1.1 Film adhesive supported metal-to-metal

1.1.1.2.1 Film adhesive unsupported metal-to-metal

1.1.1.1.2 Film adhesive supported metal-to-cured composite

1.1.3.1.1 Film and foam adhesive supported metal-to-metal

1.1.2.1.1 Form adhesive supported metal-to-metal

1.1.1.1.3 Film adhesive supported metal-to-composite prepreg

1.1.4.1.1 Adhesive paste/liquid single part metal-to-metal

Table 7.2: Comparison of alternative data to 1.1.1.1.1

In the PCODEs, the first digit represents material, so this should never be changed. The second digit represents adhesives and shouldn't be changed because the other types of materials do not come in supported and unsupported forms. This means the only PCODEs that can be used when looking for surrogate data for 1.1.1.1.4 are in the form 1.1.X.X.X.

A program can be developed in EXCEL to create Table 7.2. A column is designated for each digit of the PCODE. This allows for a comparison of the values for each digit location of the PCODE and thereby generates the last column. For example, digits one and two are always similar. For Rows 2, 3, and 6, the third digit is equivalent to the third digit in row 1 (the PCODE being investigated). For Rows 3, 4, 5, and 6, the fourth digit is equivalent. For rows 2, 4, and 5, the fifth digit is equivalent.

For the "Most alike" column of Table 7.2, the PCODE the most similar to 1.1.1.1.1 is indicated by a value of 1. The increasing numbers from 1 to 5 indicate decreasing similarity. This is done in EXCEL by simply comparing the absolute value of the Z values in column 4. The smallest absolute Z value receives a 1 and increasing Z values receive consecutive numbers.

This "Most alike" column shows 1.1.2.1.1 (Row 5) is most similar to 1.1.1.1.1; therefore, the third digit is the least significant. This conclusion might not be valid if 1.1.3.1.1 (Row 4) produced a high Z value because this PCODE also has only the third digit different from 1.1.1.1.1. However, the Z value for 1.1.3.1.1 (Row 4) is also low, so the relative insignificance of the third digit is a valid conclusion. The data was not investigated for 1.1.4.1.1 because it is a different type of adhesive which does not come in a supported or unsupported form.

Since 1.1.1.1.2 (Row 3) and 1.1.1.1.3 (Row 6) are the most dissimilar to 1.1.1.1.1 (they have a high number ranking), the fifth digit is significant. Since there is only one PCODE with just the fourth digit different (1.1.1.2.1 – Row 2), it is hard to draw a conclusion about the fourth digit; however, the relative low Z value indicates the fourth digit may be somewhat insignificant.

The net conclusion from Table 2 is that the fifth digit is the most significant, followed by the fourth digit, and the third digit is relatively insignificant. The first two digits are constant. When looking for surrogate data for 1.1.1.1.4, the most similar data will be for 1.1.X.1.4, where X implies any value. If there is no data of this format, than data for 1.1.1.X.4 should be investigated. If there is still no data, than data for 1.1.X.X.4 should be investigated. The PCODEs fitting these criteria include 1.1.2.1.4, 1.1.3.1.4, 1.1.1.2.4, 1.1.2.2.4, and 1.1.3.2.4.

The next step to determine surrogate data for 1.1.1.1.4 might be to look at the PCODE 1.1.1.2.4, which has the same last and first three digits as the PCODE 1.1.1.1.1. It is important that the investigation for significant PCODE digits is performed on at least two PCODEs (here 1.1.1.1.1 and 1.1.1.2.4) that are similar to the PCODE for which the data is desired but unpopulated (1.1.1.1.4). If the comparison is only performed for one PCODE, the conclusions may not be valid.

PCODE						Average	Std dev	Z (30 samples)	Most Alike = 1.1.1.2.4	Similar digits
1	1	1	2	4		1.04	0.02			
1	1	1	2	1		1.06	0.02	-3.872983346	5	1, 2, 3, 4
1	1	2	2	4		1.045	0.04	-0.612372436	1	1, 2, 4, 5
1	1	1	2	2		1.05	0.013	-2.296172409	3	1, 2, 3, 4
1	1	1	2	3		1.01	0.033	4.258283122	6	1, 2, 3, 4
1	1	4	2	4		1.03	0.009	2.497399895	4	1, 2, 4, 5
1	1	3	2	4		1.03	0.04	1.224744871	2	1, 2, 4, 5

1.1.1.2.4 Film adhesive unsupported cured-to-cured composite

1.1.1.2.1 Film adhesive unsupported metal-to-metal

1.1.2.2.4 Foam adhesive unsupported cured-to-cured composite

1.1.1.2.2 Film adhesive unsupported metal-to-cured composite

1.1.1.2.3 Film adhesive unsupported metal-to-composite prepreg

1.1.1.2.4 Film adhesive unsupported cured-to-cured composite

1.1.4.2.4 Paste/liquid adhesive two part cured-to-cured composite

1.1.3.2.4 Film and foam adhesive unsupported cured-to-cured composite

Table 7.3: Comparison of alternative data to 1.1.1.2.4

The PCODEs 1.1.1.3.X, 1.1.1.4.X, 1.1.1.2.5, are not options because these possibilities don't exist (they are infeasible).

Table 7.3 shows 1.1.2.2.4 (Row 3) is the most similar to 1.1.1.2.4 and 1.1.3.2.4 (Row 7) and 1.1.4.2.4 (Row 6) are also quite similar to 1.1.1.2.4, so again the third digit is the least significant. It also shows the fifth digit is again most significant because 1.1.1.2.3 (Row 5), and 1.1.1.2.1 (Row 2) are most dissimilar to 1.1.2.2.4. Here the Z value for 1.1.1.2.2 (Row 4) is less than –2.5; however, since the Z value is still larger than that for 1.1.3.2.4 and 1.1.2.2.4, the conclusion that the fifth digit is significant can still be considered correct. Row 4 can be excluded from the analysis.

The results of Table 7.3 agree with the results for Table 7.2 – the fifth digit is the least significant, and the third digit is relatively insignificant. Therefore, the options for surrogate data, in decreasing order of applicability, are 1.1.3.1.4, 1.1.2.1.4, 1.1.4.1.4, 1.1.2.2.4, and 1.1.3.2.4. In looking for surrogate data for 1.1.1.1.4, it might also be appropriate to investigate the PCODE 1.1.1.1 but this may cause some of the aggregate data problems discussed in Section 5.4.

7.5 Continually changing classification scheme

The large aerospace company is currently using various versions of classification schemes in different departments because their hierarchies change so often. A standardized hierarchy is needed.

7.6 Conclusion

This chapter addressed the two technical barriers to design usage of PCD. A method for developing a consistent classification hierarchy for PCDBs was proposed. This consistent hierarchy simplifies the process of choosing surrogate data for unpopulated indexes. A quantitative method for determining surrogate data is proposed. The problem of allowing designers to choose infeasible indexes was also addressed in this chapter.

8 Conclusion

This chapter first explains the contributions of this work. It then details further research that can be performed to continue the concepts developed in this thesis.

8.1 Contributions

This thesis provides an analysis of current industry usage of process capability databases. It shows that the assumption of process capability data access and its use by design is not valid. Both the consulted companies and the academic literature on design variation reduction tools show that design use of PCD is desired and needed.

This thesis explains how PCD is accessed by design, how the data progresses between the various functions, and the database hierarchy for a large aerospace company. Several suggested improvements for these systems are also provided. A framework for the steps required to obtain correct PCD was outlined. The various desired uses for PCD by design pending the elimination of all of the current barriers were also detailed.

This thesis identifies the ten organizational and technical barriers to PCD usage by design. The three organizational issues were identified by the survey of numerous industries: lack of a company-wide vision of PCD usage, lack of trust between suppliers and customers, and lack of communication between functional groups within an organization.

The seven technical issues were alluded to in the survey results and were shown in the case study of a large aerospace company. This thesis proposes methods to overcome all seven of these technical issues:

- First, the software prototype presents the process capability graphically as a series of runs.
- Second, the theory and examples for enabling PCD uncertainty to be quantified were provided. This uncertainty was displayed as a confidence interval in the prototype software.

- Third, the hierarchy of the large aerospace company was analyzed and methods for making it more consistent were proposed.
- Fourth, details on how infeasible indexes could be eliminated using the prototype software were provided.
- Fifth, quantative methods for determining surrogate data for unpopulated indexes were detailed and examples of their usage were provided.
- Sixth, methods of visualizing the samples of data that make up the aggregate data were provided. Issues of groupings of data and outlier data are discussed.
- Seventh, quantitative methods for determining if two runs can be combined were detailed with examples. Visual methods for determining outlier data were also explained.

8.2 Further research

Many ideas were generated for work that could be done to continue the work started in this thesis. PCDB hierarchies, which are discussed in Chapter 6 need to be improved and standardized. More quantitative methods should be developed for determining appropriate aggregate and surrogate data. Potential future enhancements to the prototype software would also contribute to the process capability database field. These should include: only allowing the user to choose parameters for which data is included in the database, allowing the plotted data to be averaged, and allowing the multiple values to be chosen for each parameter.

The first possibility for future work revolves around populating PCDBs. It should be determined if particular data is more important that other data. This could be determined at the large aerospace company by examining what type of data are most frequently requsted in PCARs. This information is useful because, if certain data is needed more than others, this data should be the first data to be populated into the databases. Methodologies for populating databases efficiently could also be investigated.

Alternative quantitative methods to the Z value for determining surrogate data for unpopulated indexes and combinable data for statistical validity and aggregate data should be determined.

Another area that could be pursued is determining the statistical validity of aggregate data. An improved hierarchy for the database should be designed such that parallel data is compatible as a potential surrogate and for combination as aggregate data.

It is currently unclear how to deal with parts that are produced through multiple processes since the current hierarchy allows for only one process to be selected. Generally, as many processes as possible are made on one machine, so the machine used for the most important process will determine the part family of the product. This is another area for potential future work. Chapter 7 shows that a better system of process capability indexing is required.

Finally, companies indicated that they would like to see cost factors integrated with the PCDBs. Ideally, cost factors could be applied to the processes such that it is possible to perform an analysis of cost factor vs. capability. The cost factors could be used when the designer is trying to chose between options by looking at the aggregate data and could be used to compare surrogate data and its monetary risk. Figure 8.1 shows what this cost vs. process capability analysis might look like.

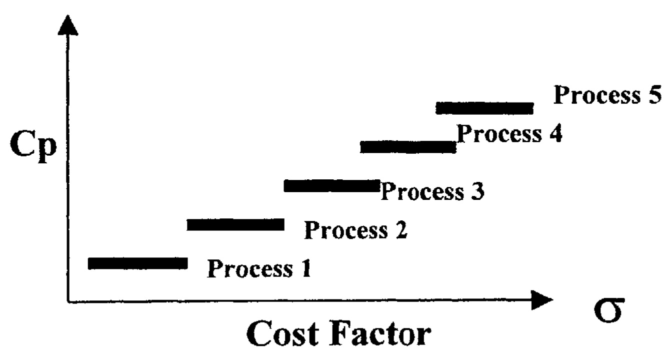


Figure 8.1: Cost factor vs. capability analysis

8.3 *Software Enhancements*

There are several features that could be added to the prototype software system to enhance its capabilities. Section 8.3.1 describes some minor modifications that could quickly be added to the prototype software. Section 8.3.2 details how outlier runs could be excluded and Section 8.3.3 how infeasible options could be eliminated. Finally, section 8.3.4 explains how the specification limits could be generated based on capability.

8.3.1 Minor modifications

The first feature would be the allowance for sorting by multiple PCODEs, LSLs, and USLs. This can be modeled after how the program currently sorts by multiple machines. Another feature would be allowing the user to choose what data he/she would like to be included in the table beside the plot. For example, if operator, date, etc. were collected in the database, these values could easily be added to the table.

A third feature would be to allow the user to go from the data plot back to the user interface to modify some of the options that he/she has selected. Currently, the program will only run one plot and then it terminates. Ideally, the user would be able to keep returning to the user interface form to select new parameters and keep plotting these parameters until he/she finds the data that he/she needs. A fourth feature would be the allowance for the input of a confidence interval value rather than always using 95%. A fifth issue is determining how this prototype software system can easily be linked to other programs such as computer aided design and analysis software.

8.3.2 Excluding outlier runs

The next feature that could be added to the software prototype is allowing the user to click on a run on the plot to exclude that run. Then there could be a command button called "Average". When this button is pressed another line would be added to both the table and the plot. This line would show the average mean shift for all of the runs that were not specifically excluded by the user. This line would show the average value for all the included data including C_{pk}, number of samples, lower confidence interval, and upper confidence interval.

8.3.3 Eliminating infeasible options

Another potential enhancement to the prototype software system is listing only available options. This completely eliminates the problem of infeasible or unpopulated indexes being chosen. For all of the lists on form 1, a designated order would need to be specified. This order might be PCODE, Machine, minimum target, maximum target, LSL, and finally USL. The user would first input a PCODE or a set of PCODEs. After this value is entered a new database would be created that contains only the data for this PCODE. This new database would be used to list only those values for machine, minimum target, maximum target, USL, and LSL that are available for that PCODE. Next, the user would chose a machine or a set of machines. Another new database would be created that contains only the data for those machines and PCODEs selected. This database would be used to create the lists of options for the targets and specification limits. After a minimum target value is chosen, the only options that should be available for the maximum target are those target values that are greater than or equal to this minimum target value. For the specification limits, each LSL can only be paired with particular USL values, so only these values should be listed.

8.3.4 Generation of specification limits based on capability

Another enhancement would be allowing the user to choose to obtain either C_{pk} by entering specification limits or to obtain specification limits by inputting a value for C_{pk}. The formulas that are used for LSL, USL, and C_{pk} show that they are only dependent on each other and on the average value and standard deviation. The enhancement of obtaining the average value for all of the plotted data for the standard deviation and the mean shift was already discussed. These values for the mean shift the standard deviation together with the values for either C_{pk} or LSL/USL could be used to obtain the value desired. Details on this enhancement are provided in Appendix F.

The large aerospace company would like to use this prototype software system to determine the validity of the data in their process capability database. The prototype software can used to determine outlier values to determine if there were problems during data collection. The software can also determine problems by comparing the runs for particular parameters that should be similar to determine if they actually are.

Many companies have requested that a case study be performed to determine the benefits of using PCD in the design process. The idea is to have a designer design a part without using PCD and then design a similar part using PCD and specifically identify where the benefits lie. It is possible that there would be realized benefits in the time for the initial design, in design for manufacturability, in a better quality part, and/or in the part being easier or faster to manufacture. Another case study that has been requested is to find a company that makes good use of their PCD in design and determine what enables them to do this.

The results of this research clearly indicate the compelling value of using process capability for product development and show that industry currently isn't taking advantage of this. By diagnosing the hindrances to use of PCD by design and potential solutions to them, this thesis is a launch pad for the numerous follow-on actions presented. Industry and academia alike can use the results and framework provided in this thesis to develop more detailed case studies and analyses.

References

Andersson, P. (1994) "A Semi-Analytic Approach to Robust Design in the Conceptual Design Phase." *Research in Engineering Design - Theory Applications and Concurrent Engineering* **8**, pp. 229-239.

ANSI Y14. 5M (1994) "Dimensioning and Tolerancing."

Baldwin, R. A. and M. J. Chung (1995) "Managing Engineering Data for Complex Products." *Research in Engineering Design - Theory Applications and Concurrent Engineering* 7(4), pp. 215-231.

Batchelor, R. and K.G. Swift (1996) "Conformability analysis support of design for quality." *ImechE Journal of Engineering Manufacture*. **210**, pp. 37-47.

Campbell, R. I. and M. R. N. Bernie (1996) "Creating a Database of Rapid Prototyping System Capabilities." *Journal of Materials Processing Technology* **61**(1-2), pp. 163-167.

Chase, K. W., J. S. Gao, S. P. Magleby and C. D. Sorensen (1996) "Including Geometric Feature Variations in Tolerance Analysis of Mechanical Assemblies." *IIE Transactions* **28**(10), pp. 795-807.

Chen, S. -L. and K. -J. Chung (1996) "Selection of the Optimal Precision Level and Target Value for a Production Process: The Lower-Specification-Limit Case." *IIE Transactions* **28**, pp. 979-985.

Clausing, Don. "Reusability in Product Development". *Engineering Design Conference 1998.* Uxbridge, England.

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

Devore, Jay L. (1987) *Probability and Statistics for Engineering and the Sciences Second Edition.* Brooks/Cole Publishing Company. Monterey, CA.

Ertan, Basak. (1998) *Systematic Organizational Inhibitors to Variation Risk Management at Boeing C-17 Airlift and Tanker Division.* LAI Case Study.

Fowlkes, William Y. and Clyde M. Creveling. (1995) *Engineering Methods for Robust Product Design.* Addison-Wesley Publishing Company. Reading, MA.

Frey, D., K. Otto and J. Wysocki (1998) "Evaluating Process Capability Given Multiple Acceptance Criteria." *Journal of Manufacturing Science and Engineering* (to appear).

Gaither, Norman. (1994) *Production and Operations Management.* Harcourt Brace Company. Orlando, FL.

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

Ingle, Timothy. (1997) *Leveraging the Learning Process in Manufacturing*. MSc, MIT, Cambridge.

Kalpakjian, Serope. (1995) *Manufacturing Engineering and Technology*. Addison-Wesley Publishing Company. Reading, MA.

Lee, D.J. and A.C. Thornton. (1996) "The Identification and Use of Key Characteristics in the Product Development Process". *ASME Design and Theory Methodology Conference.* Irvine, CA.

Lee, S. J., B. J. Gilmore and M. M. Ogot (1993) "Dimensional Tolerance Allocation of Stochastic Dynamic Mechanical Systems Through Performance and Sensitivity Analysis." *ASME Journal of Mechanical Design* **115**(3), pp. 392-402.

Leland, Cheryl. (1997) *A Cultural Analysis of Key Characteristic Selection and Team Problem Solving during an Automobile Launch.* S.M., MIT, Cambridge.

Lin, C. Y., W. H. Huang, M. C. Jeng and J. L. Doong (1997) "Study of an Assembly Tolerance Allocation Model Based Monte Carlo Simulation." *Journal of Materials Processing Technology* **70**(1-3), pp. 9-16.

Liu, C. S., S. J. Hu and T. C. Woo (1996) "Tolerance Analysis for Sheet Metal Assemblies." *ASME Journal of Mechanical Design* **118**(1), pp. 62-67.

Lucca, A. M., K. N. Berti and D. C. Cerveny (1995) "Statistical Tolerance Allocation in design utilizing historical supplier process data." *Advances in Electronic Packaging*, ASME.

Michelena, N. F. and A. M. Agogino (1994) "Formal Solution of N-Type Taguchi Parameter Design Problems with Stochastic Noise Factors." *ASME Journal of Mechanical Design* Vol. **116**(2), pp. 501-507.

Nagler, Greg. (1996) *Sustaining Competitive Advantage in Product Development: A DFM Tool for Printed Circuit Assembly*. MSc, MIT, Cambridge.

Naish, J. C. (1996) "Process Capability Modeling in an Integrated Concurrent Engineering System -- The feature-oriented Capability Module." *Journal of Materials Processing Technology* **61**, pp. 124-129.

Noltemeyer, T. (1994) "Process Centering - the Taguchi Loss Function". Ceramic Engineering Science Processing **15**(3), pp. 91-96.

Owen, Jean. (1998) "Job Shops: On the Free Highway Called the Net, Small Shops Can Travel as Fast as the Big Guys". *Manufacturing Engineering.* pp. 80-89.

Parkinson, A., Variability (1995) "Robust Mechanical Design Using Engineering Models." *Transactions of the ASME, Journal of Mechanical Design* **117**, pp. 48-54.

Parkinson, A., C. Sorensen and N. Pourhassan (1993) "A General Approach for Robust Optimal Design." *Transactions of the ASME: Journal of Mechanical Design* pp. 75-81.

Perzyk, M. and O. K. Meftah (1998) "Selection of Manufacturing Process in Mechanical Design." *Journal of Materials Processing Technology* **76**(1-3), pp. 198-202.

Phadke, M. S. (1989) *Quality Engineering Using Robust Design.* PTR Prentice-Hall Inc, Englewood Cliffs, New Jersey.

Srinivasan, R. S ., K. L. Wood and D. A. McAdams (1996) "Functional Tolerancing: A Design for Manufacturing Methodology." *Research in Engineering Design - Theory Applications and Concurrent Engineering* **8**(2), pp. 99-115.

Srinivasan, V. and M.A. O'Connor (1994) "On Interpreting Statistical Tolerancing." *Manufacturing Review* **7**, pp. 304-311.

Tata, M. and A. Thornton (1999) "Process Capability Database Usage in Industry: Myth vs. Reality." *Design for Manufacturing Conference, ASME Design Technical Conferences,* Las Vegas, NV.

Thornton, A.C. (1998) "Optimism vs. Pessimism: Design Decisions in the Face of Process Capability Uncertainty". *Journal of Mechanical Design, Transactions of the ASME*, (submitted September 1998).

Thornton, A. C. (1999) "Variation Risk Management Using Modeling and Simulation." *ASME Journal of Mechanical Design, Transactions of the ASME* (accepted Feb. 1999).

Thornton, A., S. Donnelly, B. Ertan (working paper) "More Than Just Robust Design: Why Product Development Organizations Still Contend with Variation and Its Impact on Quality".

Ting, K. -L. and Y. Long (1996) "Performance Quality and Tolerance Sensitivity of Mechanisms." *ASME Journal of Mechanical Design* **118**(1), pp. 144-150.

Ulrich, Karl T. and Stephen D. Eppinger, (1995) *Product Design and Development*, McGraw-Hill, New York, New York.

Wang, N. and T. M. Ozsoy (1993) "Automatic Generation of Tolerance Chains from Mating Relations Represented in Assembly Models." *ASME Journal of Mechanical Design* **115**(4), pp. 757-761.

Zhang, Z. and X.D. Fang (1996) "Fit Capability Indices and Their Applications" *International Journal of Production Research* **34**(11) pp. 3079-3094.

Zhang, C. C. and H. P. Ben Wang (1998) "Robust Design of Assembly and Machining Tolerance Allocations." *IIE Transactions* **30**(1), pp. 17-29.

Appendix A: Process Capability Database Survey Questions

Process Capability Database Questionnaire

Name ______________________________
Company ____________________________
Division _____________________________
Location: City _______________________ State ______
Title _______________________________
Phone Number ________________________
Email Address _________________________________

1. Does your company have a process capability database?
__________ YES __________ NO

2. What is your interaction level with process capability databases and process capability data? PLEASE CHECK ALL THAT APPLY.
____ use process capability data to work on part drawings
____ populate databases with process capability data
____ maintain process capability databases
____ other ____
____ other ____

3. Which area is your work most closely related to? PLEASE CHECK ALL THAT APPLY.
_____ Manufacturing
_____ Design
_____ Quality
_____ Other, please specify ____

4. Is your process capability database set up for: PLEASE CHECK ALL THAT APPLY.
________ a particular division, _________________________ (specific division)
________ several divisions, ____________________________ (specific divisions)
________ a specific plant, _____________________________ (specific plant)
________ several plants, ______________________________ (specific plants)
________ the entire company
________ other, please specify ______________________________________

5. For these survey questions, will you be basing your answers on the process capability database that you have for:
_____ your particular site
_____ your particular division
_____ your entire company
_____ other ____
_____ other ____

6. How long has your company been using this database?
_________ hasn't been used yet
_________ 0 – 6 months
_________ 6 months – 1 year
_________ 1 year – 3 years
_________ 3 years – 10 years
_________ over 10 years

QUESTIONS 7 - 29 APPLY TO PROCESS CAPABILITY DATABASES FOR PARTS MANUFACTURED INTERNALLY ONLY (i.e. NOT SUPPLIER PARTS).

7. Why was the internal part database developed? PLEASE CHECK ALL THAT APPLY.
_________ process monitoring
_________ design feedback
_________ inspection
_________ regulatory requirements
_________ other, please specify ____
_________ other, please specify ____

8. What software do you use for the internal part database?
Title ____________________________ Manufacturer __________________________
Title ____________________________ Manufacturer __________________________

9. What information is contained in the internal part database? PLEASE CHECK ALL THAT APPLY.
_______ raw part data
_______ raw key characteristic data
_______ SPC data (C_{pk} values)
_______ part drawings
_______ other, please specify ____
_______ other, please specify ____

10. What is the information on the internal part database used for? PLEASE CHECK ALL THAT APPLY.
_______ monitoring parts as they are being processed to make sure that they meet specifications
_______ designing new parts with more appropriate tolerances
_______ designing parts more quickly based on similar older part designs
_______ other, please specify ____
_______ other, please specify ____

11. How does the internal part data get entered into the database? PLEASE CHECK ALL THAT APPLY.
_______ manually into a computer on the shop floor
_______ manually into computer in a designated area not on the shop floor
_______ automatically through special program ___________________ (which program)
_______ by email
_______ paper recordings in file cabinet
_______ other, please specify ____
_______ other, please specify ____

12. Who enters the internal part data into the database? PLEASE CHECK ALL THAT APPLY.
_______ the machine operators
_______ the assembly operators
_______ the quality control people
_______ other, please specify ____
_______ other, please specify ____

13. How often is the internal part data entered into the database? PLEASE CHECK ALL THAT APPLY.
_______ data for every n^{th} part entered for every batch of parts
_______ data entered only for parts that have not been manufactured before
_______ data entered for the first batch of the part every time production of that particular part begins again
_______ other, please specify ____
_______ other, please specify ____

14. Once the internal part data is entered, how often is it updated for people with access to look at it? PLEASE CHECK ALL THAT APPLY.
_______ information available immediately once it's entered
_______ updated data available every hour
_______ updated data available every shift
_______ updated data available every day
_______ updated data available every week
_______ updated data available every month
_______ other, please specify ____

15. Is access to the internal part database:
________ available to all company employees and all suppliers
________ available to all company employees
________ limited access to ___________________________ (which groups/functions)

16. Why doesn't everyone have access to the internal part data? PLEASE CHECK ALL THAT APPLY.
______ everyone doesn't need data
______ some groups/functions aren't trained on how to access the data
______ confidential information that can't be provided to suppliers
______ other, please specify ____
______ other, please specify ____

17. How is the internal part data accessed by the people who use the data? PLEASE CHECK ALL THAT APPLY.
_______ internet
_______ intranet
_______ shop floor computers
_______ forms requesting particular data to be provided by group who knows how to use database
_______ other, please specify
_______ other, please specify

18. Who uses the internal part data that your company/division has in your database? PLEASE CHECK ALL THAT APPLY.

___ design engineers at your site/division
___ design engineers across the entire company
___ manufacturing engineers at your site/division
___ manufacturing engineers across the entire company
___ quality engineers at your site/division
___ quality engineers across the entire company
___ other ____
___ other ____

19. When the internal part data is accessed, how is the information indexed or what information must the user input into the system in order to find the appropriate data? PLEASE CHECK ALL THAT APPLY.

_______by part number
_______ by feature number
_______ by manufacturing process
_______ by key characteristic number
_______ by feature type
_______ by machine used to make part/feature
_______ other, please specify ______________________________________
_______ other, please specify ______________________________________

20. In what format does your internal part database present the data? PLEASE CHECK ALL THAT APPLY.

____ raw data
____ control charts
____ histograms
____ other ____
____ other ____

21. What other systems are linked to your internal part process capability database? PLEASE CHECK ALL THAT APPLY.

_______a part drawing system, title ______________________________________
_______ a measurement system, title ______________________________________
_______ a design system, title ____
_______ other, please specify ____
_______ other, please specify ____

22. What programs are currently linked to your internal part database? PLEASE CHECK ALL THAT APPLY.

_____ variation simulation analysis
_____ design of experiments
_____ computer aided design
_____ other ____
_____ other ____

23. How do you use your internal part data in other programs such as CAD, VSA, etc? PLEASE CHECK ALL THAT APPLY.

____ direct link between systems
____ copy data into other systems
____ other ____
____ other ____

24. What percentage of your internal parts are contained in the database?

_______ 0-10% _______ 11-20% _______ 21-30%
_______ 31-40% _______ 41-50% _______ 51-60%
_______ 61-70% _______ 71-80% _______ 81-90%
_______ 91-100%

25. If your internal part database is populated with data for only some parts, which parts have data? PLEASE CHECK ALL THAT APPLY.

_____ parts manufactured most recently
_____ parts for which data entry is automated
_____ parts that are the most expensive
_____ parts where tolerances are the most critical
_____ parts that contain no data already
_____ parts that have undergone a process improvement
_____ parts for which data would be needed most frequently
_____ newest parts
_____ oldest parts
_____ other ____
_____ other ____

26. Why is your internal part database not populated with data for all of your internally manufactured parts? PLEASE CHECK ALL THAT APPLY.

____ don't have the people resources to populate the database
____ don't have the financial resources to populate the database
____ database is new, so only parts manufactured recently have data
____ data not being used, so no incentive to populate database
____ other ____
____ other ____

27. When you add internal part data to areas of your database that are already populated, what do you do with the data?

___ it is averaged in with the old data
___ it is keep separate from the old data with its date label; however, it is possible to average the old and new data
___ other ____
___ other ____

28. How do you record process improvements or problems in your internal part database? PLEASE CHECK ALL THAT APPLY.

____ separate from database
____ notes linked to particular PCD
____ don't record
____ data for particular process, material, feature, etc separate for each process improvement
____ data not collected when there is a problem
____ eliminate all old PCD for that particular process, material, feature, etc when process improvement is made
____ other ____
____ other ____

29. What percentage of internal parts and assemblies at your company/division are designed/toleranced using process capability databases?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

QUESTIONS 30-48 APPLY TO PROCESS CAPABILITY DATABASES FOR PARTS PROVIDED BY SUPPLIERS ONLY.

30. Do you have a process capability database for supplier parts?
_______ YES - **GO TO QUESTION 32** _______ NO

31. If NO, why?
____
____
SKIP TO QUESTION 49

32. Why was the supplier database developed? PLEASE CHECK ALL THAT APPLY.
_________ process monitoring
_________ design feedback
_________ inspection
_________ regulatory requirements
_________ other, please specify _______________________________
_________ other, please specify _______________________________

33. What is the information in the supplier database used for? PLEASE CHECK ALL THAT APPLY.
_______ choosing between several suppliers for a new part
_______ designing parts more quickly based on what parts are available from suppliers
_______ other, please specify _______________________________
_______ other, please specify _______________________________

34. In what form do you receive the data from the supplier? PLEASE CHECK ALL THAT APPLY.
_______ handwritten on paper
_______ saved on disk in a spreadsheet
_______ saved on a disk in a process capability database program
_______ other, please specify _______________________________
_______ other, please specify _______________________________

35. Do you request this supplier data in a particular program, and if so, which?
_______ YES _______ NO
Title ______________________ Manufacturer ______________________
Title ______________________ Manufacturer ______________________

36. What type of process capability data do you ***require*** from suppliers? PLEASE CHECK ALL THAT APPLY.
_______ raw part data
_______ raw key characteristic data
_______ SPC data (C_{pk} values)
_______ part drawings
_______ none
_______ other, please specify ____
_______ other, please specify ____

37. What type of process capability data do you ***obtain*** from the supplier? PLEASE CHECK ALL THAT APPLY.
_______ raw part data
_______ raw key characteristic data
_______ SPC data (C_{pk} values)
_______ part drawings
_______ other, please specify ____
_______ other, please specify ____

38. Is this supplier database:
_______ a separate database than the one for parts manufactured internally
_______ the same database as the one for parts manufactured internally
_______ other, please specify ____

39. How does the supplier data get entered into the database? PLEASE CHECK ALL THAT APPLY.
_______ manually by the supplier
_______ manually by some group/function in your organization whom obtains the data
from the supplier, __________________________ (which group)
_______ automatically through special program _________________ (which program)
_______ by email
_______ paper recordings in file cabinet
_______ other, please specify ____
_______ other, please specify ____

40. How often is the supplier data entered into the database? PLEASE CHECK ALL THAT APPLY.
_______ data for every n^{th} part entered for every batch of parts
_______ data entered only for parts that have not been manufactured before
_______ data entered for the first batch of the part every time production of that particular
part begins again
_______ other, please specify ____
_______ other, please specify ____

41. Once the supplier data is entered, how often is it updated for people with access to look at it?
_______ information available immediately once it's entered
_______ updated data available every hour
_______ updated data available every shift
_______ updated data available every day
_______ updated data available every week
_______ updated data available every month
_______ other, please specify ____

42. Is access to the supplier database:

________ available to all company employees and all suppliers
________ available to all company employees
________ limited access to ______________________________ (which groups/functions)
________ other, please specify ____

43. Why doesn't everyone have access to the supplier data? PLEASE CHECK ALL THAT APPLY.

______ everyone doesn't need data
______ some groups/ functions aren't trained on how to access the data
______ confidential information that can't be provided to suppliers
______ other, please specify ____
______ other, please specify ____

44. How is the supplier data accessed by the people who use the data? PLEASE CHECK ALL THAT APPLY.

_______ internet
_______ intranet
_______ shop floor computers
_______ forms requesting particular data to be provided by group who knows how to use database
_______ other, please specify ____
_______ other, please specify ____

45. When the supplier data is accessed, how is the information indexed or what information must the user input into the system in order to find the appropriate data? PLEASE CHECK ALL THAT APPLY.

_______by part number
_______ by feature number
_______ by manufacturing process
_______ by key characteristic number
_______ by feature type
_______ by machine used to make part/feature
_______ other, please specify ____
_______ other, please specify ____

46. What percentage of your supplier parts are contained in the database?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

47. What other systems are linked to your supplier process capability database? PLEASE CHECK ALL THAT APPLY.

_______a part drawing system, title ____
_______ a measurement system, title _______________________________________
_______ a design system, title ____
_______ other, please specify ____
_______ other, please specify ____

48. What percentage of your internal parts are designed/toleranced using supplier PCD?

_____ 0%
_____ 1-10%
_____ 11-20%
_____ 21-30%
_____ 31-40%
_____ 41-50%
_____ 51-60%
_____ 61-70%
_____ 71-80%
_____ 81-90%
_____ 91-100%

QUESTIONS 49 – 64 PERTAIN TO HOW THE INTERNAL AND/OR SUPPLIER PART PROCESS CAPABILITY DATABASE IS USED OR HOW IT COULD BE USED BY DESIGN.

49. Is your process capability database used for design?

_______ YES _______ NO – GO TO QUESTION 50

50. How do you use your process capability database for design?

____

____

51. Why don't you use your process capability databases for design?

____

____

52. How would you like to use your process capability database for design?

____

____

53. What information do designers want in process capability databases (i.e. what information do they need to improve the design process)? PLEASE CHECK ALL THAT APPLY.

_____ Cp and Cpk
_____ Pareto charts
_____ cause & effect diagrams
_____ feature/part/process spoilage history
_____ results from gage R & R
_____ standard deviation
_____ raw data
_____ target costs
_____ mean shifts
_____ Xbar from target
_____ Yield
_____ Special causes (i.e. process improvements, process problems)
_____ machine
_____ operator
_____ date
_____ control charts
_____ other ____
_____ other ____

54. How do designers currently obtain process capability information for creating new designs? PLEASE CHECK ALL THAT APPLY.

_____ word of mouth from manufacturing
_____ PCDBs directly
_____ request for information to be gathered from PCDB by someone else
_____ don't use PCD at all
_____ reference manuals
_____ other ____
_____ other ____

55. How do designers at your company/division use process capability data? PLEASE CHECK ALL THAT APPLY.

____ allocate tolerances based on machine capability
____ allocate tolerances based on tolerances for similar older part designs
____ input into variation simulation analysis
____ choose between options for new part designs
____ other ____
____ other ____

56. Why is your process capability database not utilized fully in the product development process? PLEASE CHECK ALL THAT APPLY.

____ because the software systems to use PCD data is not integrated or does not exist
____ because the data structures are difficult to search
____ because there are no design incentives to use data
____ because there is not a clear understanding of customer satisfaction
____ because there is a lack of management support
____ because there is a lack of clear communication and cooperation between functions
____ other ____
_____ other ____

57. What percentage of the time do designers at your company use variation simulation analysis to allocate tolerances on your designs?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

58. What percentage of the time do designers at your company use robust design to allocate tolerances on your designs?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

59. What percentage of the tolerances that designers at your company specify are set based on real process capability data (i.e. data that has been collected and is presented in some type of printed form as opposed to data from someone based on experience rather than on recordings)?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

60. What percentage of the tolerances that designers at your company specify are set based on guesses about capability by designers?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

60. What percentage of the tolerances that designers at your company specify are set based on manufacturing expert knowledge?

_____ 0%	_____ 31-40%	_____ 71-80%
_____ 1-10%	_____ 41-50%	_____ 81-90%
_____ 11-20%	_____ 51-60%	_____ 91-100%
_____ 21-30%	_____ 61-70%	

61. Does your company have any proof that it is beneficial for design to use process capability data?

____ YES ____ NO

If yes, what ?

____

____

____

62. Would your company/division be willing to participate in such an experiment?

____ YES ____ NO

63. Assuming the databases are fully populated rank the incentives that would prompt designers using the process capability data?

_____ management requirement to use data

_____ management monetary incentives to use data

_____ case study showing that designs are made more manufacturable by using process capability data

_____ short amount of time required to obtain data How long? ______________

_____ other ____

_____ other ____

QUESTIONS 64 – 78 ARE GENERAL QUESTIONS ABOUT BOTH INTERNAL AND SUPPLIER PART PROCESS CAPABILITY DATABASES.

64. Do you have any methods in place to determine how frequently the process capability data is utilized?
____ YES ____ NO

65. Do you have any methods in place to determine by whom the process capability data is utilized?
____ YES ____ NO

66. Do you have any methods in place to determine for what the process capability data is utilized?
____ YES ____ NO

67. If yes to questions 64, 65, or 66 what is/are the methods?

68. What short term improvements are currently planned for your process capability database(s)?

69. What long term improvements are currently planned for your process capability database?

70. What is your ideal future vision for your process capability database in terms of the type and quantity of information that it contains and in terms of what it is utilized for?

71. What parts of your process capability database are in greatest need of improvement? Please rank with 1 being most important and higher numbers being less important.
____ user interface
____ population of data
____ accessibility to entire database
____ accuracy of data
____ usage of data
____ hierarchy of data so that it is easier to find quickly
____ other ____
____ other ____

72. Have resources for developing your process capability database been increased or decreased during the past year?
_____ increased _____ decreased

Why?
____
____
____

73. Has your company/division had any significant successes in using process capability databases for design or other areas?
_____ YES _____ NO

If yes, what?
____
____
____
____

74. Do you have any strategies for improving PCDBs that might be useful to other companies? If so, what?
____
____
____
____
____

75. What further information would you like to have on process capability databases in various industries?
____
____

76. Can I contact you for follow-up information? How should I contact you?
____
____

77. Who are the other good contacts on this topic at your company and how can I contact them?
____
____

78. Are you willing to let me summarize the data (i.e. no specifics) to send back out as a pre-paper summary?
_______ YES _______ NO

COMMENTS ABOUT THIS:

____

____

____

ANY ADDITIONAL COMMENTS:

____

____

____

____

____

____

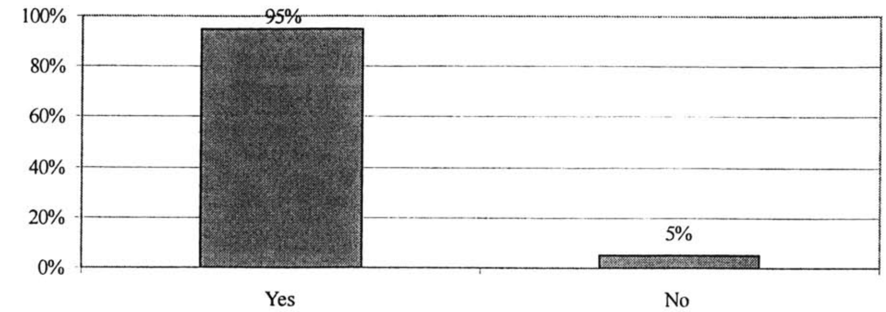
____

Appendix B: Questionnaire 1 and 2 Responses

1. Does your company have a process capability database?

Yes	*95%*
No	*5%*

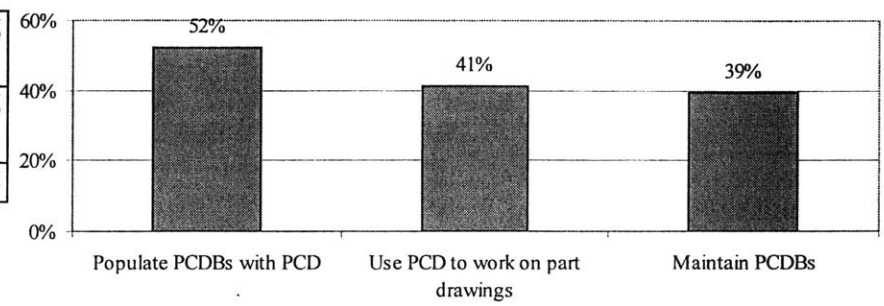
No. respondents = 42



2. What is your interaction level with process capability databases and process capability data?

Use PCD to work on part drawings	*41%*
Populate PCDBs with PCD	*52%*
Maintain PCDBs	*39%*

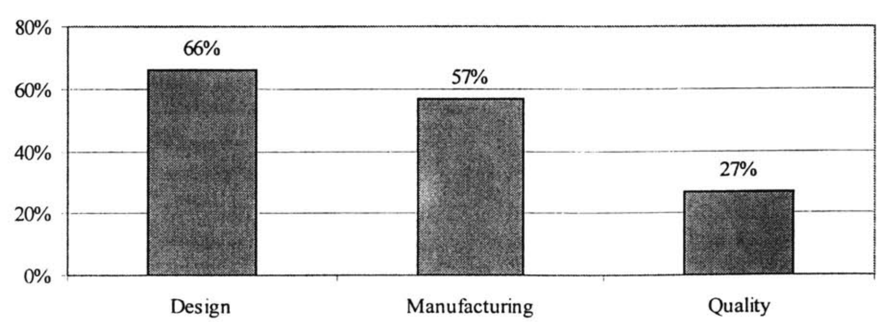
No. respondents = 18



3. Which area is your work most closely related to?

Manufacturing	*57%*
Design	*66%*
Quality	*27%*

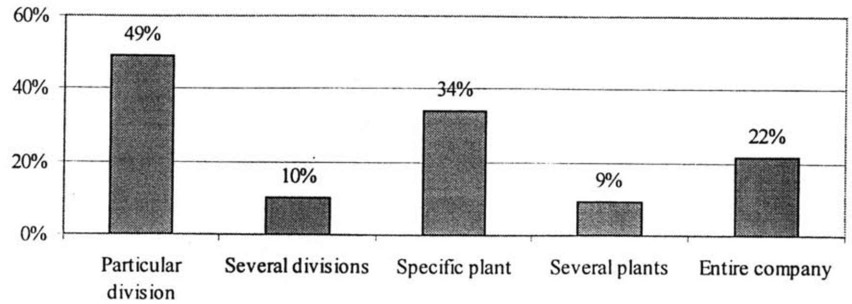
No. respondents = 19



4. Is your process capability database set up for:

Particular division	*49%*
Several divisions	*10%*
Specific plant	*34%*
Several plants	*9%*
Entire company	*22%*

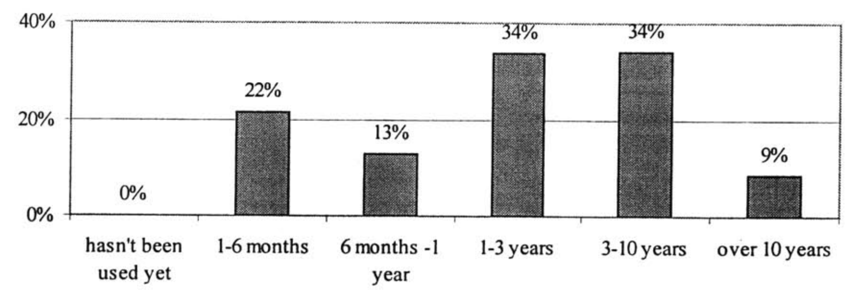
No. respondents = 36



5. How long has your company been using this database?

hasn't been used yet	*0%*
1-6 months	*22%*
6 months -1 year	*13%*
1-3 years	*34%*
3-10 years	*34%*
over 10 years	*9%*

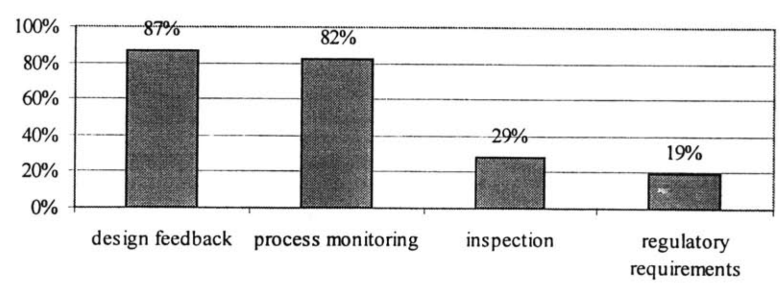
No. respondents = 37



6. Why was the internal part database developed?

process monitoring	*82%*
design feedback	*87%*
inspection	*29%*
regulatory requirements	*19%*

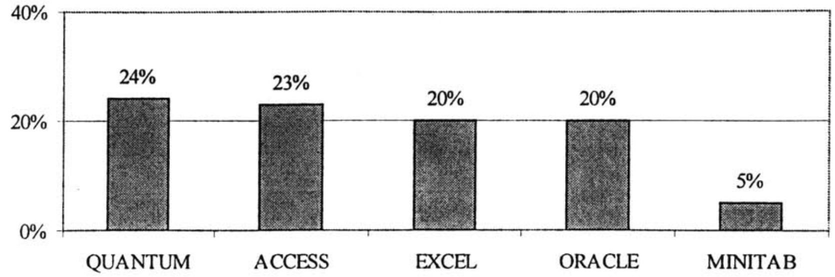
No. respondents = 35



7. What software do you use for the internal part database?

MINITAB	*5%*
ACCESS	*23%*
EXCEL	*20%*
ORACLE	*20%*
QUANTUM	*24%*

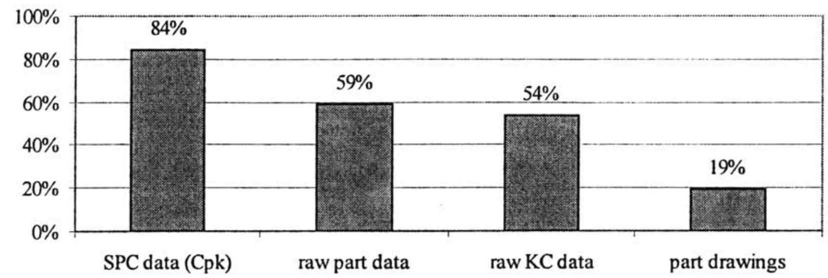
No. respondents = 32



8. What information is contained in the internal part database?

raw part data	*59%*
raw KC data	*54%*
SPC data (Cpk)	*84%*
part drawings	*19%*

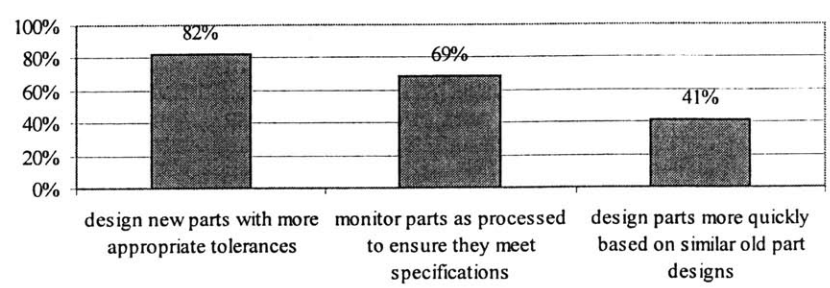
No. respondents = 34



9. What is the information on the internal part database used for?

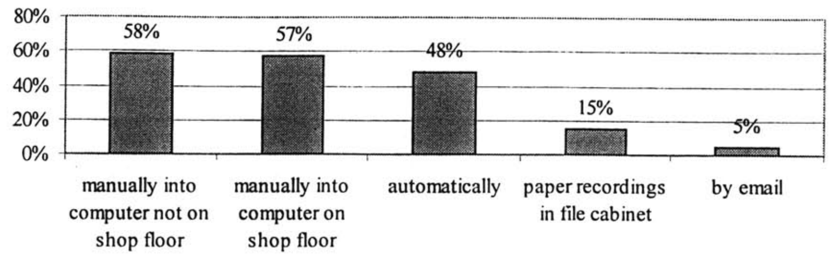
No. respondents = 34

Monitor parts as processed to ensure they meet specifications	*69%*
design new parts with more appropriate tolerances	*82%*
design parts more quickly based on similar old part designs	*41%*



10. How does the internal part data get entered into the database?

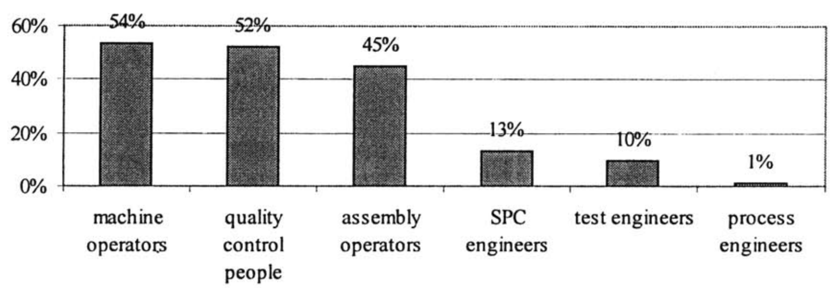
manually into computer not on shop floor	58%
manually into computer on shop floor	57%
automatically	48%
paper recordings in file cabinet	15%
by email	5%



No. respondents = 33

11. Who enters the internal part data into the database?

machine operators	54%
assembly operators	45%
quality control people	52%
process engineers	1%
test engineers	10%
SPC engineers	13%

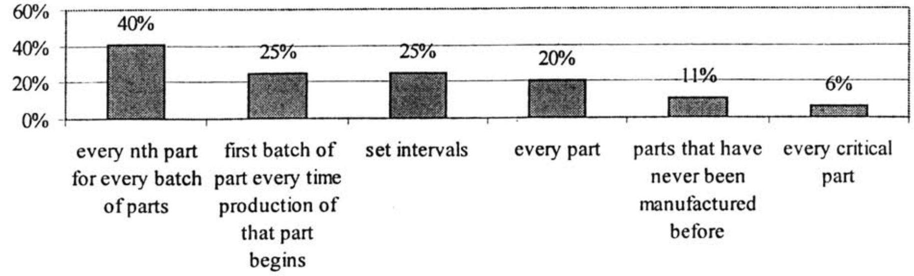


No. respondents = 32

12. How often is the internal part data entered into the database?

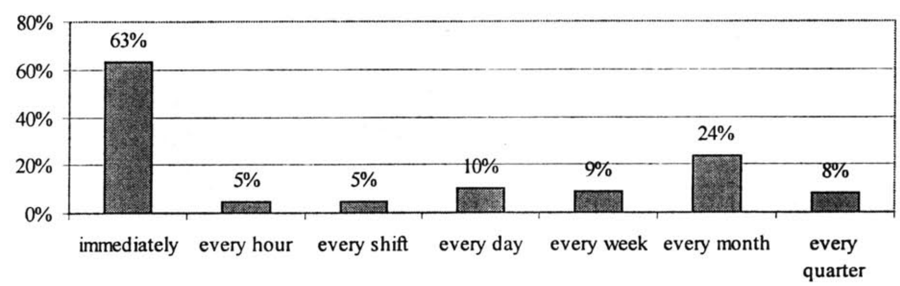
No. respondents = 29

every nth part for every batch of parts	40%
parts that have never been manufactured before	11%
first batch of part every time production of that part begins	25%
every part	20%
every critical part	6%
set intervals	25%



13. Once the internal part data is entered, how often is it updated for people with access to look at it?

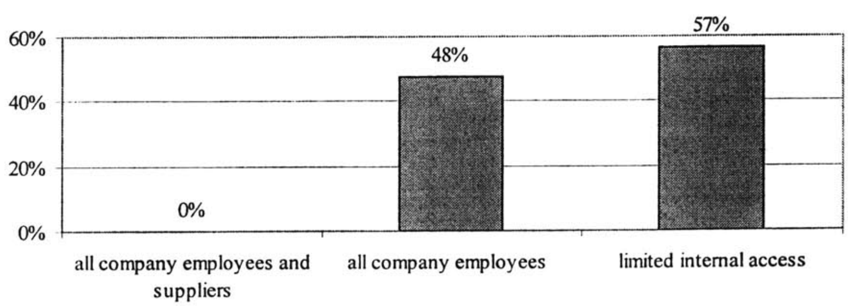
immediately	*63%*
every hour	*5%*
every shift	*5%*
every day	*10%*
every week	*9%*
every month	*24%*
every quarter	*8%*



No. respondents = 27

14. Who has access to internal databases?

all company employees and suppliers	*0%*
all company employees	*48%*
limited internal access	*57%*

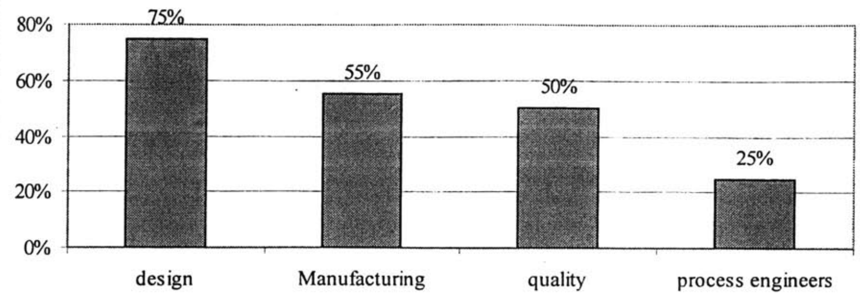


No. respondents = 34

15. Who has limited access to internal databases?

design	75%
Manufacturing	55%
quality	50%
process engineers	25%

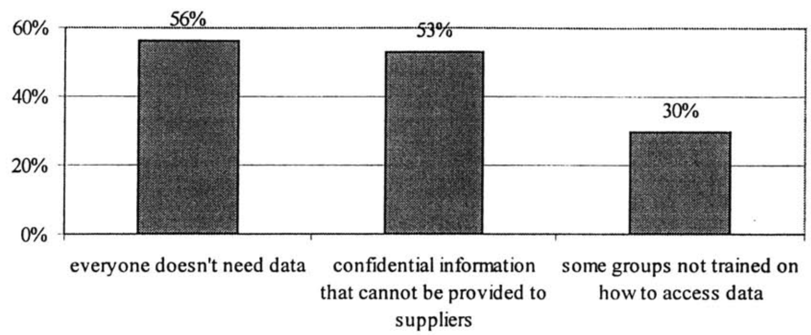
No. respondents = 18



16. Why doesn't everyone have access to the internal part data?

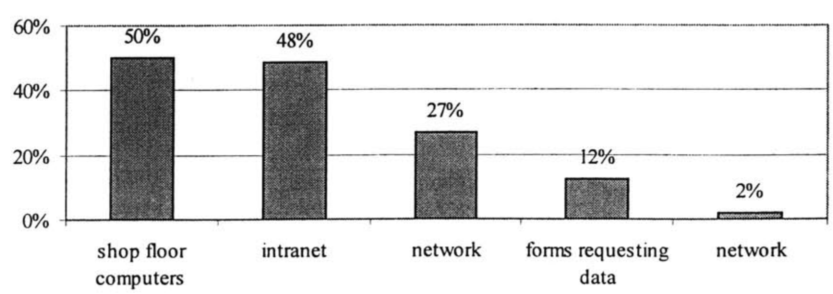
everyone doesn't need data	56%
some groups not trained on how to access data	30%
confidential information that cannot be provided to suppliers	53%

No. respondents = 31



17. How is the internal part data accessed by the people who use the data?

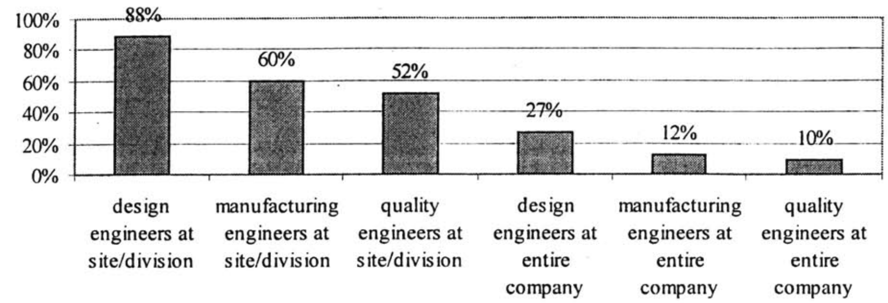
internet	2%
intranet	48%
shop floor computers	50%
forms requesting data	12%
network	27%



No. respondents = 32

18. Who uses the internal part data that your company/division has in your database?

design engineers at site/division	*88%*
design engineers at entire company	*27%*
manufacturing engineers at site/division	*60%*
manufacturing engineers at entire company	*12%*
quality engineers at site/division	*52%*
quality engineers at entire company	*10%*

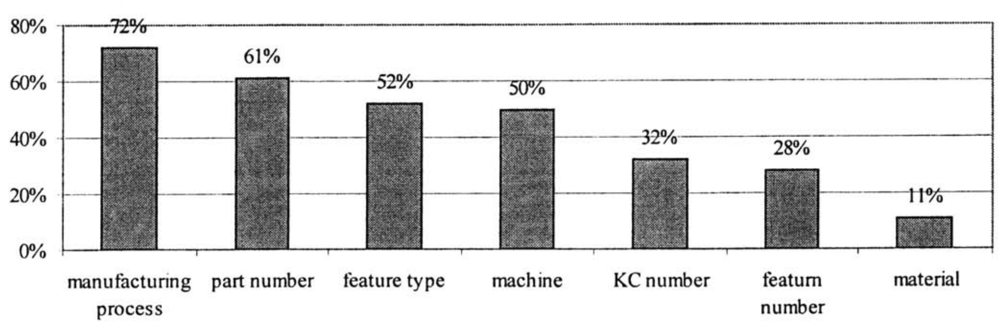


No. respondents = 18

19. When the internal part data is accessed, how is the information indexed or what information must the user input into the system in order to find the appropriate data?

part number	*61%*
feature number	*28%*
manufacturing process	*72%*
KC number	*32%*
feature type	*52%*
machine	*50%*
material	*11%*

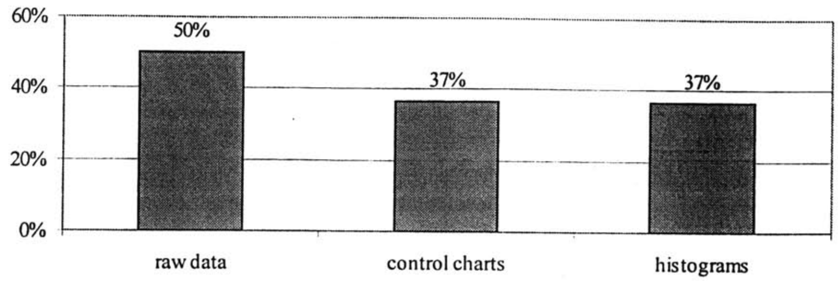
No. respondents = 33



20. In what format does your internal part database present the data?

raw data	*50%*
control charts	*37%*
histograms	*37%*

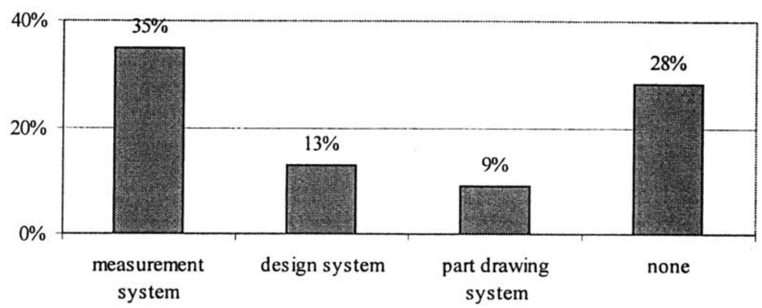
No. respondents = 14



21. What other systems are linked to your internal part process capability database?

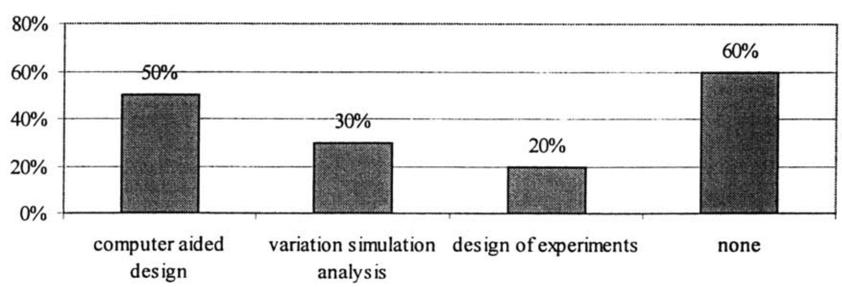
part drawing system	*9%*
measurement system	*35%*
design system	*13%*
none	*28%*

No. respondents = 26



22. What programs are currently linked to your internal part database?

variation simulation analysis	*30%*
design of experiments	*20%*
computer aided design	*50%*
none	*60%*

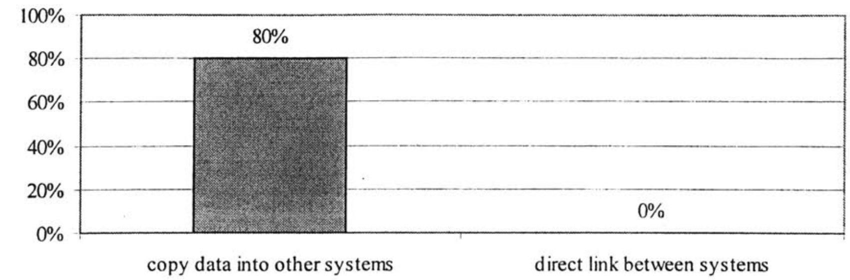


No. respondents = 12

23. How do you use your internal part data in other programs such as CAD, VSA, etc?

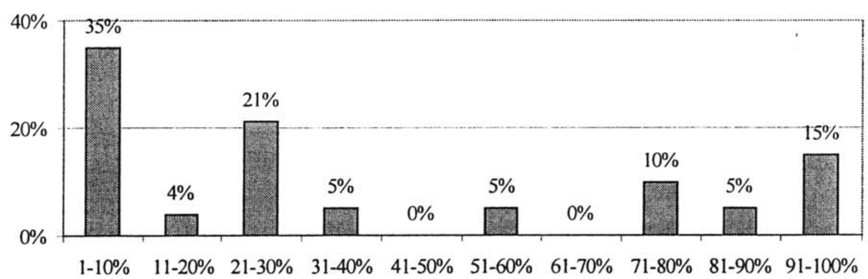
direct link between systems	*0%*
copy data into other systems	*80%*

No. respondents = 9



24. What percentage of your internal parts are contained in the database?

0%	*0%*
1-10%	*35%*
11-20%	*4%*
21-30%	*21%*
31-40%	*5%*
41-50%	*0%*
51-60%	*5%*
61-70%	*0%*
71-80%	*10%*
81-90%	*5%*
91-100%	*15%*

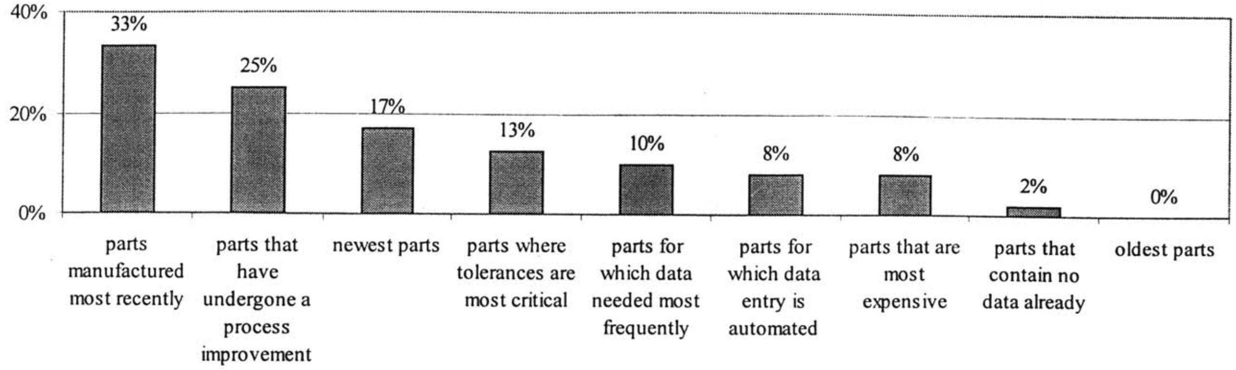


No. respondents = 31

25. If your internal part database is populated with data for only some parts, which parts have data?

parts manufactured most recently	*33%*
parts for which data entry is automated	*8%*
parts that are most expensive	*8%*
parts where tolerances are most critical	*13%*
parts that contain no data already	*2%*
parts that have undergone a process improvement	*25%*
parts for which data needed most frequently	*10%*
newest parts	*17%*
oldest parts	*0%*

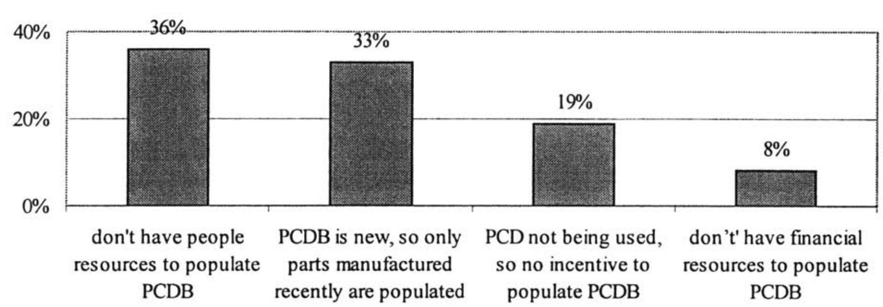
No. respondents = 14



26. Why is your internal part database not populated with data for all of your internally manufactured parts?

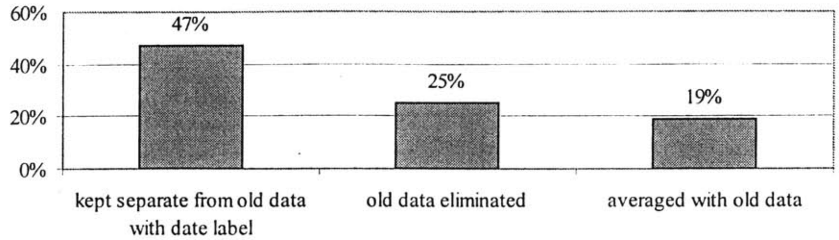
don't have people resources to populate PCDB	*36%*
don't have financial resources to populate PCDB	*8%*
PCDB is new, so only parts manufactured recently are populated	*33%*
PCD not being used, so no incentive to populate PCDB	*19%*

No. respondents = 12



27. When you add internal part data to areas of your database that are already populated, what do you do with the data?

averaged with old data	*19%*
kept separate from old data with date label	*47%*
old data eliminated	*25%*

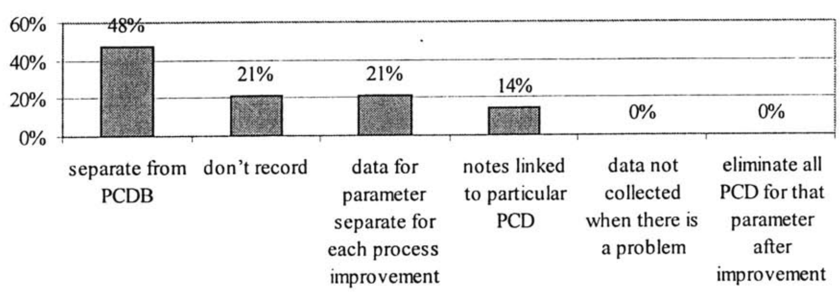


No. respondents = 14

28. How do you record process improvements or problems in your internal part database?

separate from PCDB	*48%*
notes linked to particular PCD	*14%*
don't record	*21%*
data for parameter separate for each process improvement	*21%*
data not collected when there is a problem	*0%*
eliminate all PCD for that parameter after improvement	*0%*

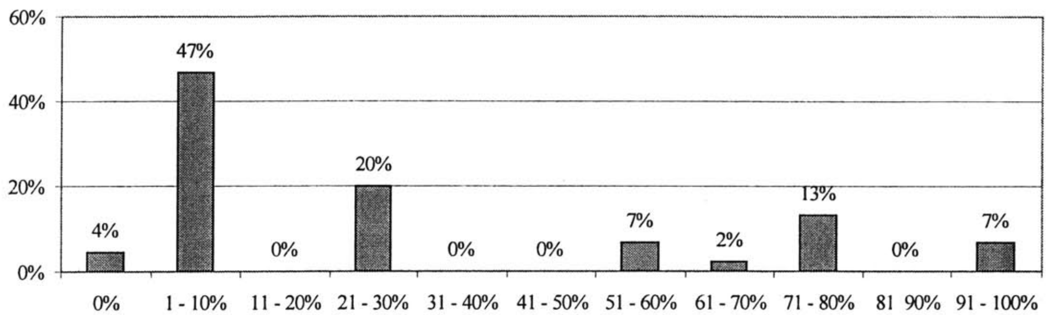
No. respondents = 15



29. What percentage of internal parts and assemblies at your company/division are designed/toleranced using process capability databases?

0%	*4%*
1 - 10%	*47%*
11 - 20%	*0%*
21 - 30%	*20%*
31 - 40%	*0%*
41 - 50%	*0%*
51 - 60%	*7%*
61 - 70%	*2%*
71 - 80%	*13%*
81 90%	*0%*
91 - 100%	*7%*

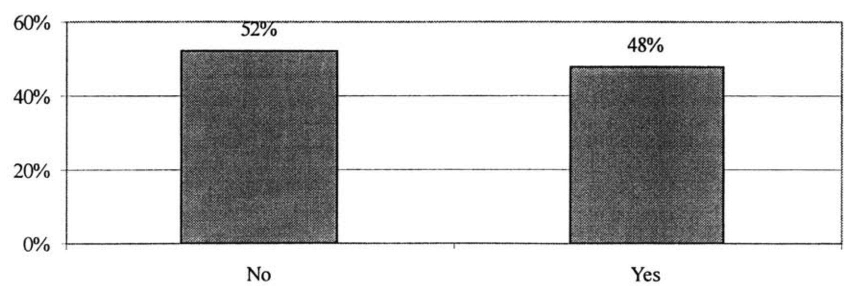
No. respondents = 14



30. Do you have a process capability database for supplier parts?

Yes	*48%*
No	*52%*

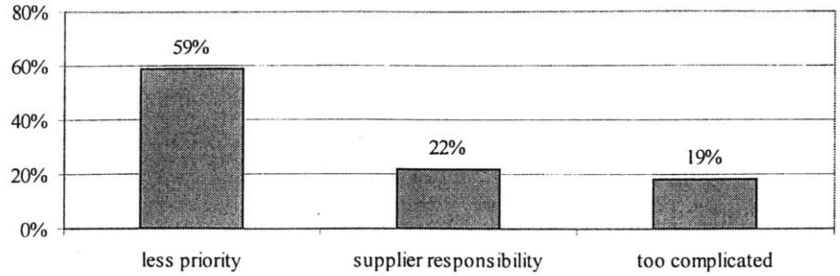
No. respondents = 36



31. If No, why?

supplier responsibility	*22%*
less priority	*59%*
too complicated	*19%*

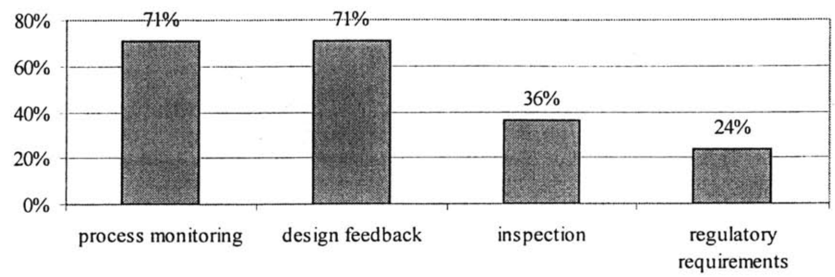
No. respondents = 14



32. Why was the supplier database developed?

process monitoring	*71%*
design feedback	*71%*
inspection	*36%*
regulatory requirements	*24%*

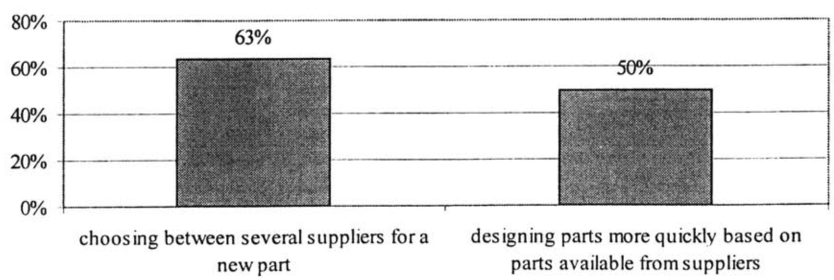
No. respondents = 17



33. What is the information in the supplier database used for?

choosing between several suppliers for a new part	*63%*
designing parts more quickly based on parts available from suppliers	*50%*

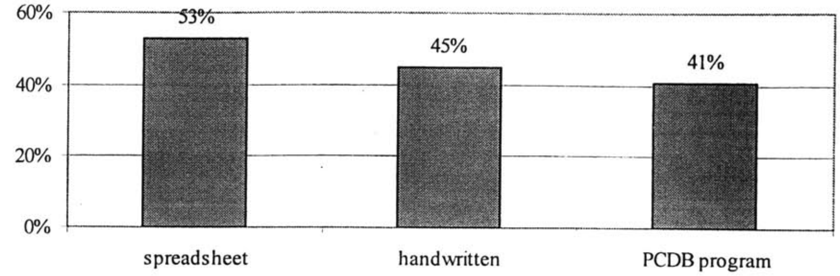
No. respondents = 15



34. In what form do you receive the data from the supplier?

handwritten	*45%*
spreadsheet	*53%*
PCDB program	*41%*

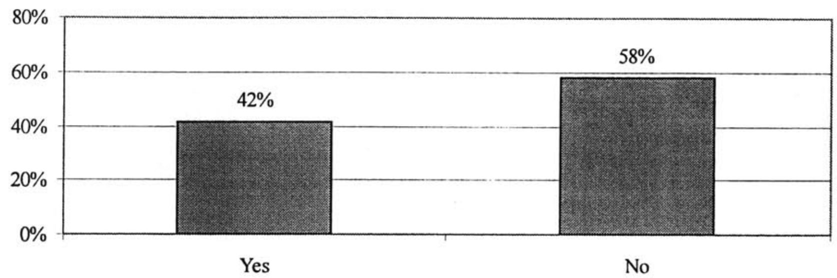
No. respondents = 16



35. Do you request this supplier data in a particular program, and if so, which?

Yes	*42%*
No	*58%*

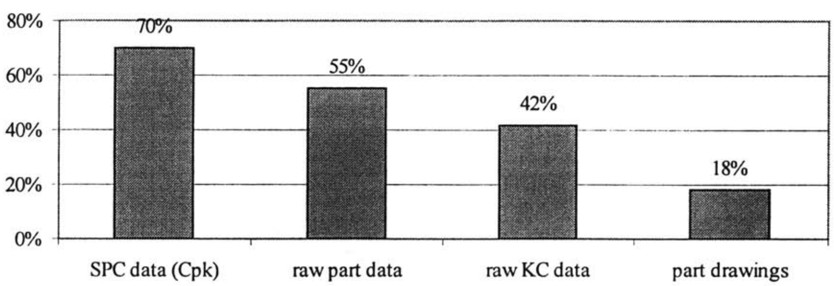
No. respondents = 17



36. What type of process capability data do you ***require*** from suppliers?

raw part data	*55%*
raw KC data	*42%*
SPC data (C_{pk})	*70%*
part drawings	*18%*

No. respondents = 16

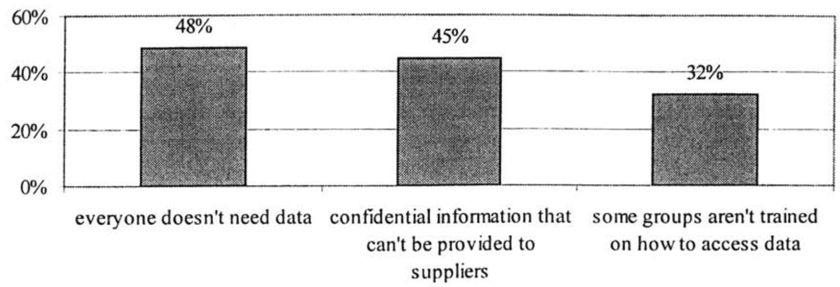


Room 14-0551
77 Massachusetts Avenue
Cambridge, MA 02139
Ph: 617.253.2800
Email: docs@mit.edu
http://libraries.mit.edu/docs

43. Why doesn't everyone have access to the supplier data?

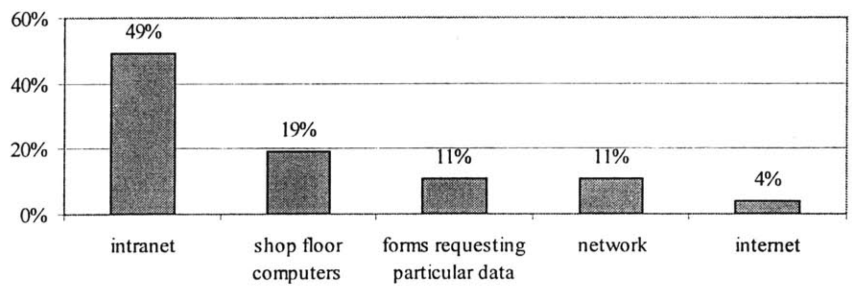
everyone doesn't need data	*48%*
some groups aren't trained on how to access data	*32%*
confidential information that can't be provided to suppliers	*45%*

No. respondents = 15



44. How is the supplier data accessed by the people who use the data?

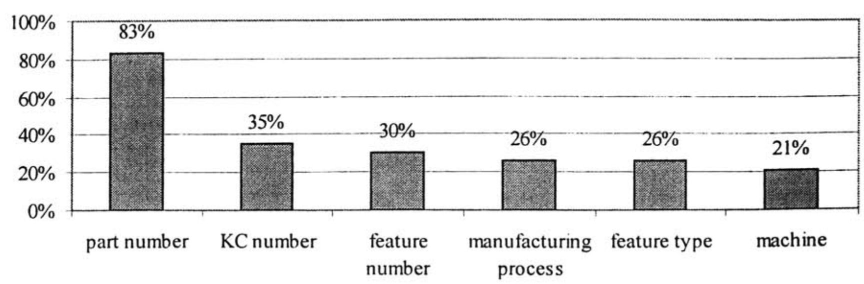
internet	*4%*
intranet	*49%*
shop floor computers	*19%*
forms requesting particular data	*11%*
network	*11%*



No. respondents = 17

45. When the supplier data is accessed, how is the information indexed or what information must the user input into the system in order to find the appropriate data?

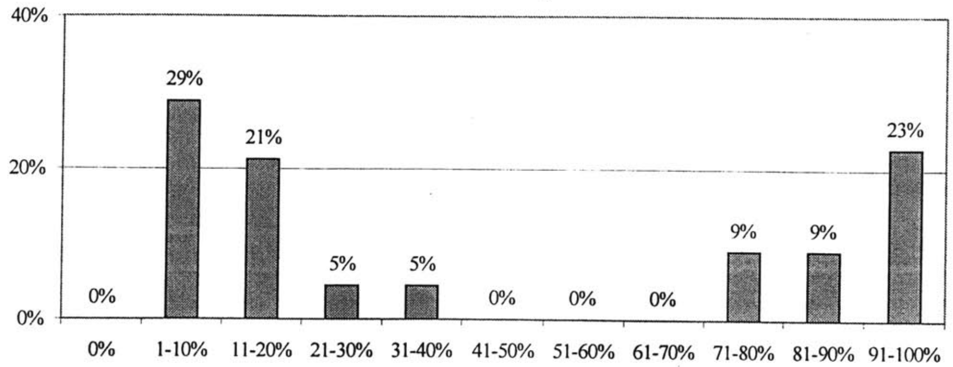
part number	*83%*
feature number	*30%*
manufacturing process	*26%*
KC number	*35%*
feature type	*26%*
machine	*21%*



No. respondents = 16

46. What percentage of your supplier parts are contained in the database?

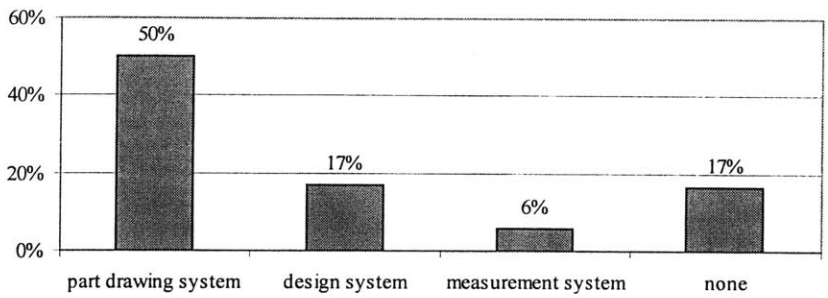
0%	*0%*
1-10%	*29%*
11-20%	*21%*
21-30%	*5%*
31-40%	*5%*
41-50%	*0%*
51-60%	*0%*
61-70%	*0%*
71-80%	*9%*
81-90%	*9%*
91-100%	*23%*



No. respondents = 16

47. What other systems are linked to your supplier process capability database?

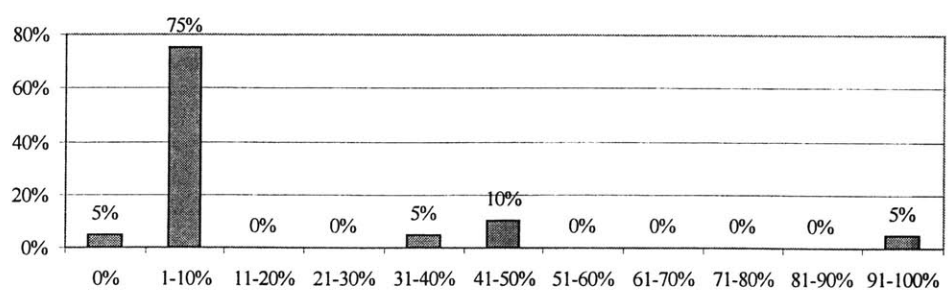
part drawing system	*50%*
measurement system	*6%*
design system	*17%*
none	*17%*



No. respondents = 7

48. What percentage of your internal parts are designed/toleranced using supplier PCD?

0%	*5%*
1-10%	*75%*
11-20%	*0%*
21-30%	*0%*
31-40%	*5%*
41-50%	*10%*
51-60%	*0%*
61-70%	*0%*
71-80%	*0%*
81-90%	*0%*
91-100%	*5%*

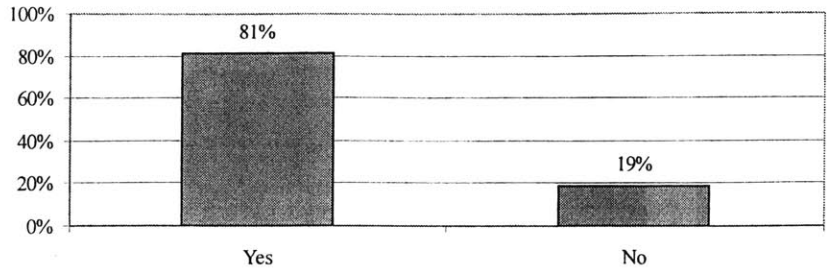


No. respondents = 21

49. Is your process capability database used for design?

Yes	*81%*
No	*19%*

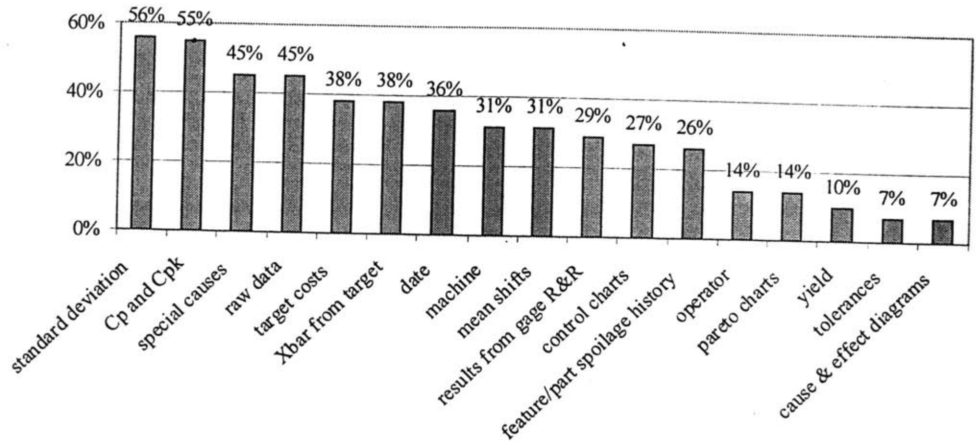
No. respondents = 35



50. What information do designers want in process capability databases (i.e. what information do they need to improve the design process)?

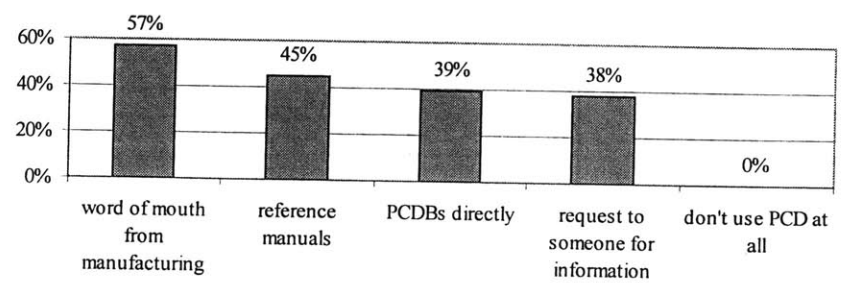
Cp and Cpk	*55%*
pareto charts	*14%*
cause & effect diagrams	*7%*
feature/part spoilage history	*26%*
results from gage R&R	*29%*
standard deviation	*56%*
raw data	*45%*
target costs	*38%*
mean shifts	*31%*
Xbar from target	*38%*
yield	*10%*
special causes	*45%*
machine	*31%*
operator	*14%*
date	*36%*
control charts	*27%*
tolerances	*7%*

No. respondents = 18



51. How do designers currently obtain process capability information for creating new designs?

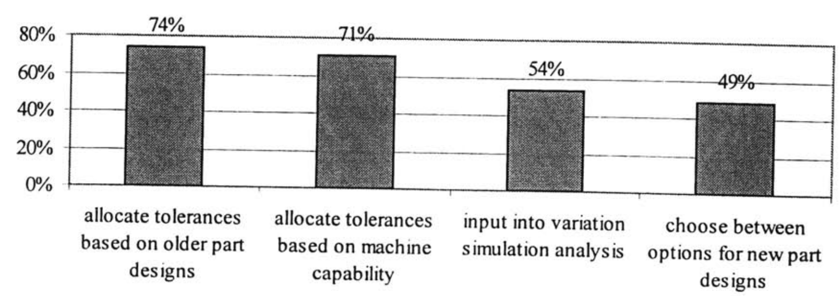
word of mouth from manufacturing	*57%*
PCDBs directly	*39%*
request to someone for information	*38%*
don't use PCD at all	*0%*
reference manuals	*45%*



52. How do designers at your company/division use process capability data?

allocate tolerances based on machine capability	*71%*
allocate tolerances based on older part designs	*74%*
input into variation simulation analysis	*54%*
choose between options for new part designs	*49%*

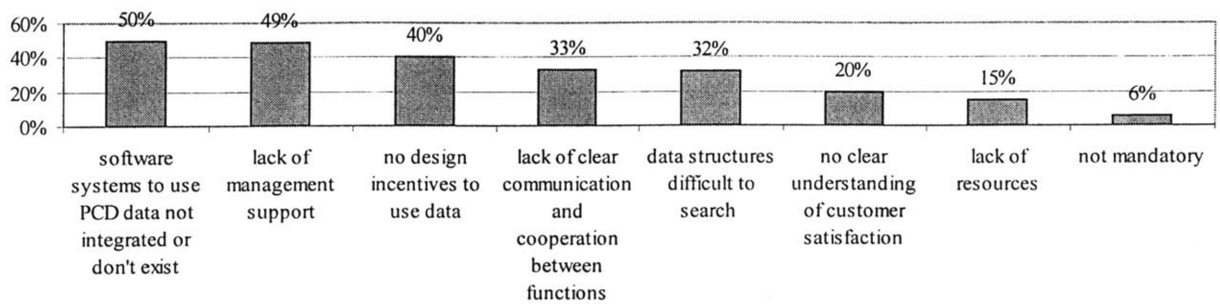
No. respondents = 17



53. Why is your process capability database not utilized fully in the product development process?

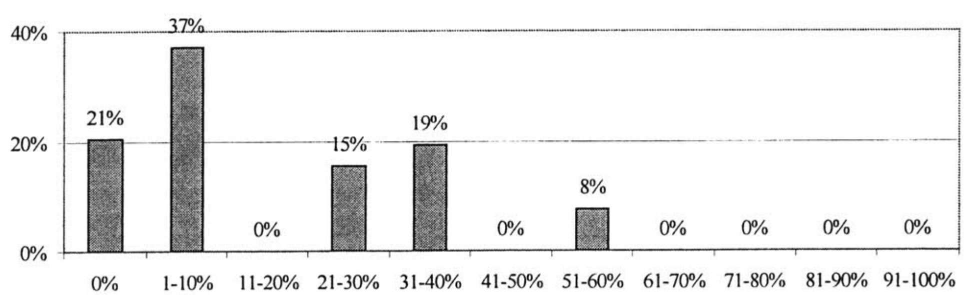
software systems to use PCD data not integrated or don't exist	*50%*
data structures difficult to search	*32%*
no design incentives to use data	*40%*
no clear understanding of customer satisfaction	*20%*
lack of management support	*49%*
lack of clear communication and cooperation between functions	*33%*
not mandatory	*6%*
lack of resources	*15%*

No. respondents = 27



54. What percentage of the time do designers at your company use variation simulation analysis to allocate tolerances on your designs?

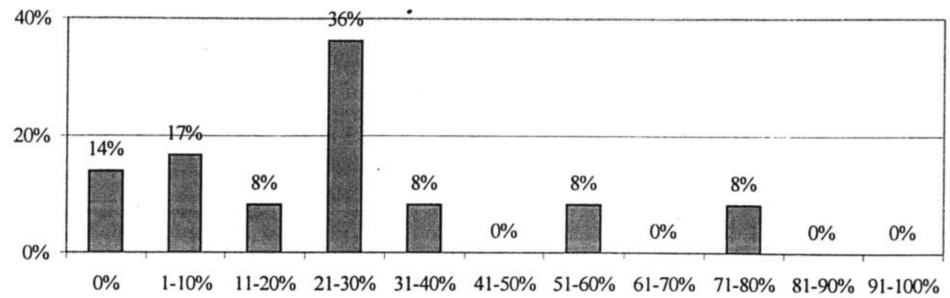
0%	*21%*
1-10%	*37%*
11-20%	*0%*
21-30%	*15%*
31-40%	*19%*
41-50%	*0%*
51-60%	*8%*
61-70%	*0%*
71-80%	*0%*
81-90%	*0%*
91-100%	*0%*



No. respondents = 17

55. What percentage of the time do designers at your company use robust design to allocate tolerances on your designs?

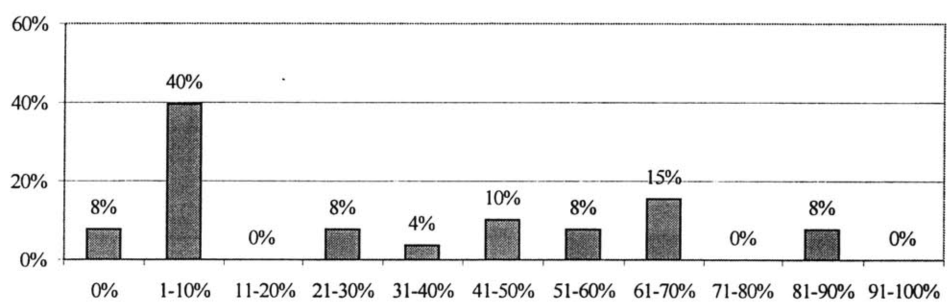
0%	*14%*
1-10%	*17%*
11-20%	*8%*
21-30%	*36%*
31-40%	*8%*
41-50%	*0%*
51-60%	*8%*
61-70%	*0%*
71-80%	*8%*
81-90%	*0%*
91-100%	*0%*



No. respondents = 15

56. What percentage of the tolerances that designers at your company specify are set based on real process capability data (i.e. data that has been collected and is presented in some type of printed form as opposed to data from someone based on experience rather than on recordings)?

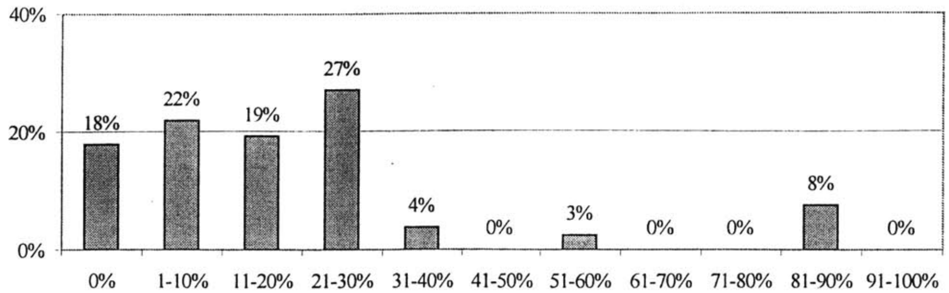
0%	*8%*
1-10%	*40%*
11-20%	*0%*
21-30%	*8%*
31-40%	*4%*
41-50%	*10%*
51-60%	*8%*
61-70%	*15%*
71-80%	*0%*
81-90%	*8%*
91-100%	*0%*



No. respondents = 17

57. What percentage of the tolerances that designers at your company specify are set based on guesses about capability by designers?

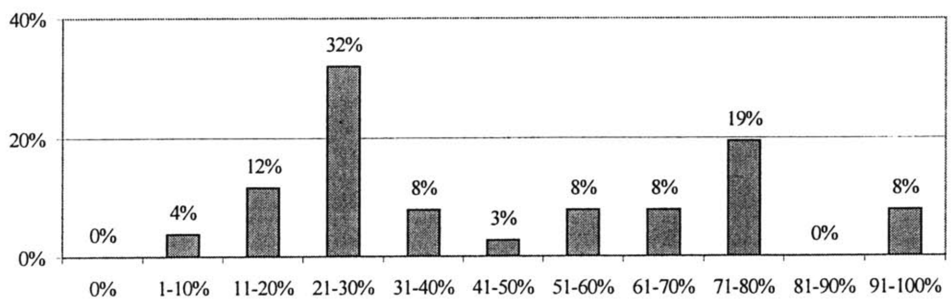
0%	*18%*
1-10%	*22%*
11-20%	*19%*
21-30%	*27%*
31-40%	*4%*
41-50%	*0%*
51-60%	*3%*
61-70%	*0%*
71-80%	*0%*
81-90%	*8%*
91-100%	*0%*



No. respondents = 17

58. What percentage of the tolerances that designers at your company specify are set based on manufacturing expert knowledge?

0%	*0%*
1-10%	*4%*
11-20%	*12%*
21-30%	*32%*
31-40%	*8%*
41-50%	*3%*
51-60%	*8%*
61-70%	*8%*
71-80%	*19%*
81-90%	*0%*
91-100%	*8%*

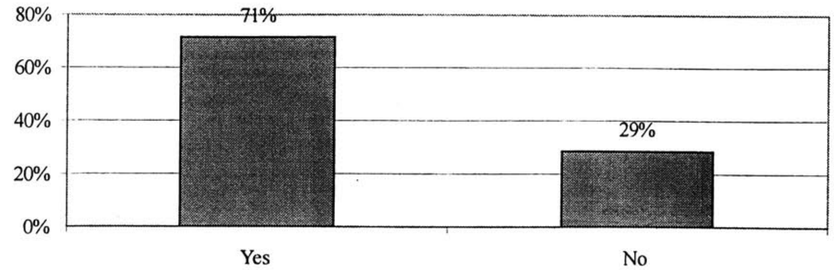


No. respondents = 17

59. Does your company have any proof that it is beneficial for design to use process capability data?

Yes	*71%*
No	*29%*

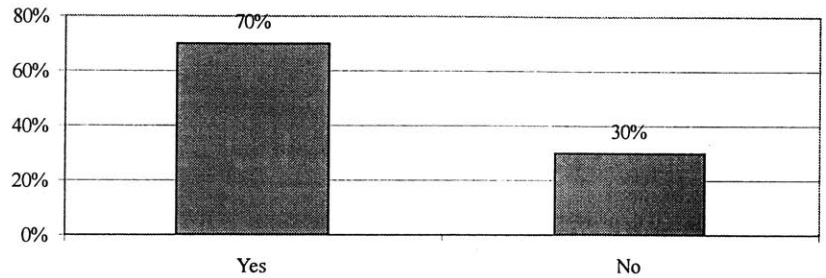
No. respondents = 17



60. Would your company/division be willing to participate in such an experiment?

Yes	*70%*
No	*30%*

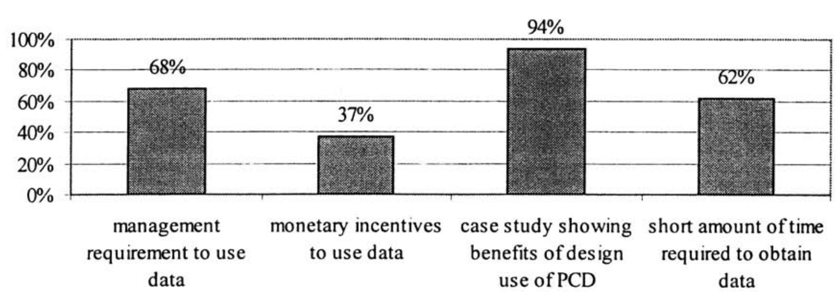
No. respondents = 12



61. Assuming the databases are fully populated rank the incentives that would prompt designers using the process capability data?

management requirement to use data	*68%*
monetary incentives to use data	*37%*
case study showing benefits of design use of PCD	*94%*
short amount of time required to obtain data	*62%*

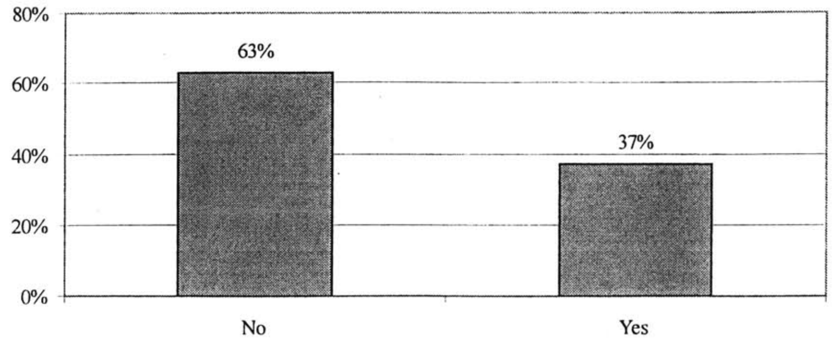
No. respondents = 15



62. Do you have any methods in place to determine how frequently the process capability data is utilized, by whom the PCD is utilized, or for what the PCD is utilized?

Yes	*37%*
No	*63%*

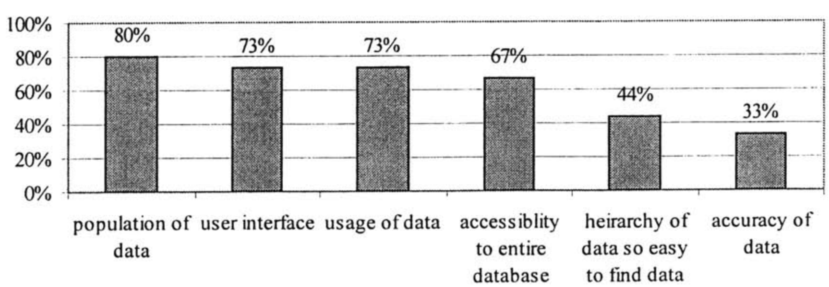
No. respondents = 28



63. What parts of your process capability database are in greatest need of improvement? Please rank with 1 being most important and higher numbers being less important.

user interface	*73%*
population of data	*80%*
accessibility to entire database	*67%*
accuracy of data	*33%*
usage of data	*73%*
hierarchy of data so easy to find data	*44%*

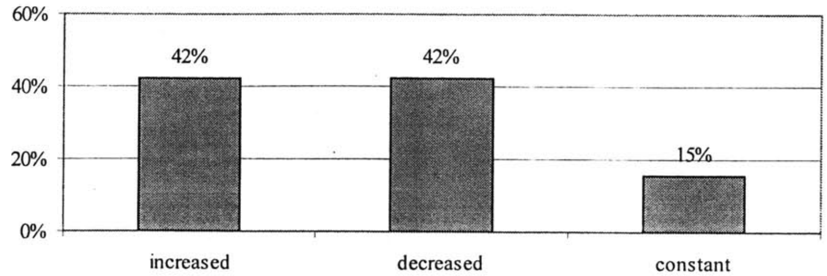
No. respondents = 15



64. Have resources for developing your process capability database been increased or decreased during the past year?

increased	*42%*
decreased	*42%*
constant	*15%*

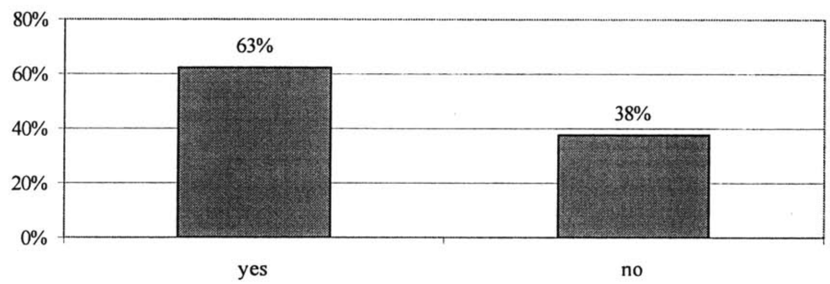
No. respondents = 15



65. Has your company/division had any significant successes in using process capability databases for design or other areas?

yes	*63%*
no	*38%*

No. respondents = 15



Appendix C: C_p and C_{pk}

C_p is "the actual observed variance divided by the print tolerance" (Noltemeyer, 1994). C_p assesses the "potential ability of the process to meet preset specification limits" (Zhang *et al.*, 1996) and does not take into account the average value of the process. On the other hand, C_{pk} provides information on how close the average value of the process is to the center of the specification limits. The formulas for both follow:

$$C_{pk} = \frac{Minimum\left(\bar{x} - LSL, USL - \bar{x}\right)}{3\sigma} \tag{1}$$

$$C_p = \frac{(USL - LSL)}{6\sigma} \tag{2}$$

Where *USL* is the upper specification limit, *LSL* is the lower specification limit, $\bar{x}$ is the average value of sample, and σ is the standard deviation of sample. The specification limits are based on design intent and the standard deviation and average are based on the process. A C_{pk} of 1.33 is equivalent to a defect rate of 30 parts per million (Batchelor *et al.*, 1996).

Figure C.1 shows how C_p and C_{pk} are different. Samples A and B have the same target, USL, and LSL, but different averages. Both normal distribution curve A and B have a C_p of 1.33, however, they have different C_{pk} values. Since the average value of sample A and the target value are equivalent, the C_{pk} of A is 1.33 or the same as the C_p because the sample is perfectly centered between its specification limits and no samples fall outside these limits. For sample B, the average value is closer to the upper specification limit then it is to the lower specification limit; therefore, the C_{pk} is some value between 0 and 1.33 as proven in Equations 3-5.

$$\frac{USL-\bar{x}_B}{3\sigma}<\frac{USL-\bar{x}_A}{3\sigma} \tag{3}$$

since $\bar{x}_A$ is located at the target,

$$\frac{USL-\bar{x}_A}{3\sigma}=\frac{USL-LSL}{6\sigma} \tag{4}$$

$$C_{pk}<C_p=1.33 \tag{5}$$

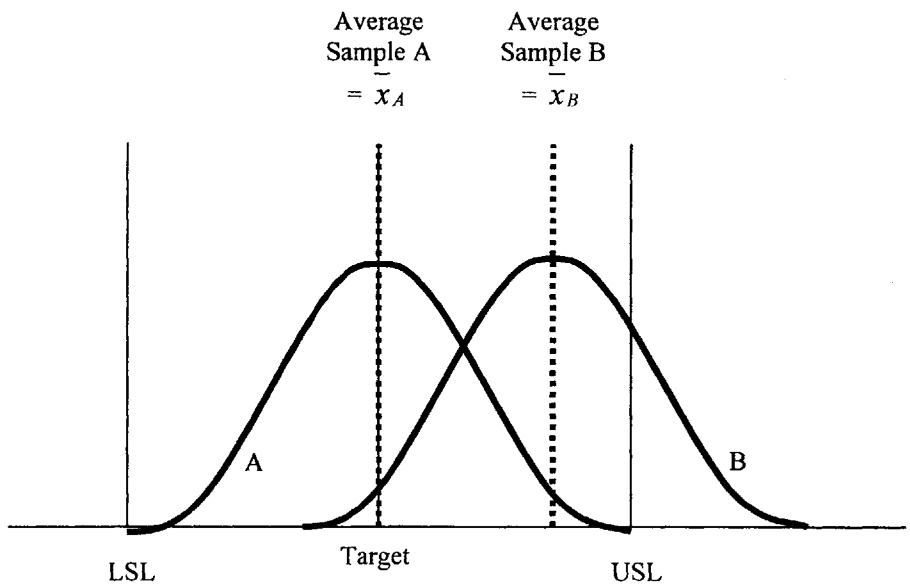


Figure C.1: Difference between C_p and C_{pk}

Appendix D: Standard Normal Curve Table

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0352	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455

Appendix E: Chi-squared Table

ν	0.975	0.025
1	0.001	5.025
2	0.051	7.378
3	0.216	9.348
4	0.484	11.143
5	0.831	12.832
6	1.237	14.44
7	1.69	4416.012
8	2.18	17.534
9	2.7	19.022
10	3.247	20.483
11	3.816	21.92
12	4.404	23.337
13	5.009	24.735
14	5.629	26.119
15	6.262	27.488
16	6.908	28.845
17	7.564	30.19
18	8.231	31.526
19	8.906	32.852
20	9.591	34.17
21	10.283	35.478
22	10.982	36.781
23	11.688	38.075
24	12.401	39.364
25	13.12	40.646
26	13.844	41.923
27	14.573	43.194
28	15.308	44.461
29	16.147	45.772
30	16.791	46.979
31	17.538	48.231
32	18.291	49.48
33	19.046	50.724
34	19.806	51.966
35	20.569	53.203
36	21.336	54.437
37	22.105	55.667
38	22.878	56.896
39	23.654	58.119
40	24.433	59.342

Appendix F: Generation of C_{pk} from Specification Limits

$$LSL = M - 3\sigma C_{pk} \qquad (1)$$

$$USL = M + 3\sigma C_{pk} \qquad (2)$$

$$C_{pk} = \frac{(M - LSL)}{3\sigma} \quad \text{OR} \quad C_{pk} = \frac{(USL - M)}{3\sigma} \quad \text{WHICHEVER IS LESS} \qquad (3)$$

where M is the mean shift, LSL is the lower specification limit, USL is the upper specification limit, σ is the standard deviation, and n is the number of samples.

For example, if the user inputs a desired C_{pk} of 1.33 and the mean shift value is 0.005 and the standard deviation value is 0.001, then the values that would be outputted for LSL and USL are:
LSL = 0.00101 OR USL = 0.00899
If the user chooses a LSL value of 0.00101, then he/she must choose a USL value of 0.0089 or greater in order to obtain a C_{pk} of 1.33. If the user choose a USL value of 0.00899, then he/she must choose a LSL value of 0.00101 or less in order to obtain a C_{pk} of 1.33.

As another example, if the user inputs LSL = 0.004 and USL = 0.012 and the mean shift value is 0.0078 and the standard deviation is 0.002, then the value that would be outputted for C_{pk} is:
C_{pk} = 0.633.