

**Continuous Process Improvement  
by Observational Studies**

*M.S. Hamada, R.J. MacKay  
and J.B. Whitney  
University of Waterloo*

May 1992

# CONTINUOUS PROCESS IMPROVEMENT BY OBSERVATIONAL STUDIES

*M. Hamada, R.J. MacKay and J.B. Whitney*

Department of Statistics and Actuarial Science  
University of Waterloo

May 1992

## *ABSTRACT*

The essential feature of an observational study performed on a process is that data are collected without deliberately changing the process. Control charting, a type of observational study, has traditionally been used to establish and maintain process stability. Once stability is achieved, the charting activity controls the process variability but does not reduce it further. This paper goes beyond control charting by proposing the use of charting techniques in conjunction with observational studies for continuous variation reduction, i.e., continuous improvement. It requires that the classification of special and common cause variation be abandoned and that instead simply sources of variation be considered. The basic strategy is to design the sampling plan to identify the largest sources of variation currently operating. As improvements are made, the sampling plans and corresponding analysis methods are updated. Consequently, most of the charting effort becomes an off-line activity. The necessity of upfront planning is discussed. It is concluded that lack of such upfront planning and purposeful sampling has resulted in the ineffective use of charting and explains a large part of the frustration that many have experienced using it.

*Key Words:* Graphical Methods, Sampling Schemes, Sources of Variation.

## Introduction

We define an observational study of a process to be one which collects data from a process without deliberately changing it. For example, control charting is a widely used type of observational study that is performed on many processes. Here we refer to charting as all activities surrounding the use of a control chart. These include planning, collecting data, analysis and making recommendations. Traditionally, charting has had several purposes:

- to provide a process record
- to signal the need for adjustment
- to establish and maintain stability.

None of these purposes is consistent with the concept of continuous improvement, i.e., variation reduction. For example, the argument has been made that once process stability is attained, only common cause variation remains in the process and that systematic changes to the process are required to achieve improvement. Such changes are considered to be outside the scope of traditional charting.

In this paper, we propose using charting to continuously reduce variation. Some immediate consequences of the proposal are that the charting has a much larger off-line component, more planning and review is required and a different terminology to describe process behavior is needed. Moreover, since this activity is beyond the domain of a local operator, the approach requires the use of a broad-based team with knowledge of the entire process under study.

Juran and Gryna (1980) discuss chronic and sporadic problems with the process. Sporadic problems result in large, immediately apparent process changes and can be detected by the traditional uses of control charts. Resolving a chronic problem, a breakthrough in Jurán's terminology, requires a different approach which is usually a change to the system. In statistical terms, a process intervention or experiment is required to achieve a breakthrough.

In this paper, we focus on the progress that can be made towards a breakthrough without such interventions in the process.

One of the key distinctions made in traditional charting is that between common (chance) and special (assignable) causes of variation. While this distinction is always made, there seems to be little agreement on definitions. For example, Deming (1986) writes,

“Shewhart used the term assignable cause of variation where I use the term special cause. I prefer the adjective special for a cause that is specific to some group of workers, or to a particular production worker, or to a specific machine, or to a specific local condition.”

and

“The term common causes for faults in the system . . . .”

Contrast this with Grant and Leavenworth (1988) who write,

“Variability of the quality characteristic may follow a chance pattern or it may behave erratically because of the occasional presence of assignable causes that can be discovered or eliminated.”

Deming seems to be separating operator controllable from other causes of variation. Grant and Leavenworth separate the causes on the basis of their detectability. This is an important point because the ability to detect a cause of variation depends heavily on the design of the data collection scheme. More recently, Pyzdek (1990) argues that the classification of causes as special or chance is detrimental because it inhibits overall variation reduction. He prefers to classify causes as visible or hidden. A visible cause has a known effect; a hidden cause is one whose effect has not yet been identified. This idea is important, but misses the point that what is visible depends on how you *view the process*. That is, a cause of variation may be visible with one data collection scheme, but hidden with another.

The treatment of causes of variation in various textbooks on statistical process control suggest that there is a confusion over how causes should be classified. Different classifications are made using criteria based on the magnitude of transmitted variation, detectability of the source of variation, control over the remedy and whether a cause has been identified or not. To avoid this confusion, we use the term “source of variation” and emphasize the relationship between the way the process is monitored (“the view of the process”) and its ability to detect the source.

MacKay (1988) suggests that sources of variation follow a Pareto principle. Consequently, a process can be continuously improved by identifying and removing the largest of the currently operating sources of variation. This strategy is iterative in that today’s relatively unimportant source of variation may become tomorrow’s largest. Note that a different sampling scheme may be required to detect this second source.

This paper concentrates on designing sampling schemes to detect sources of variation, Juran’s “diagnostic journey.” This corresponds to choosing rational subgroups in traditional charting. Once a source of variation is identified, action must be taken to reduce its effect, the “remedial” portion of the journey. Otherwise, no improvement will occur and the effort will be wasted. Concepts involving a statistical component such as moving quality upstream or Taguchi’s method for desensitizing a process to a source of variation can play an important role in the “remedial journey” but are not discussed here. Besides identifying a source of variation, the sampling plans also allow its magnitude to be evaluated. While not discussed here, its relative contribution to overall process variation can then be assessed if the overall process variation is known. In fact, before undertaking such a study, it makes sense to know what the overall process variation is, because removing an identified source of variation may result in little process variation reduction.

The paper is organized as follows. First, “charting” is embedded into a systematic problem solving approach which sets the stage for a discussion of the design of a data collection scheme. Our approach proposes a classification scheme for potential sources of variation and

employs two general principles in designing a data collection scheme. Here, we consider a simplified process of a machining operation to illustrate our approach. Next, some simple tools that can be used to analyze the collected data are discussed. The paper concludes by pointing out some strengths and limitations of the approach.

## **A Problem Solving System**

There are many versions of a step-by-step approach to problem solving (e.g., MacKay (1988)). These versions are different packaging of the following steps:

- problem selection and definition
- process description
- input identification and prioritization
- planning of data collection
- analysis
- solution generation and confirmation
- implementation and standardization.

The use of a system to solve problems is important because it enforces discipline, defines roles and responsibilities and, in the long run, is more effective. Some brief comments on the first few steps follow.

The first step is a management function and ensures that important problems are considered and that appropriate resources (people, training, time, etc.) and on-going leadership are available. The second step is the task of the problem solving team that requires a clear description of the process associated with the problem, specification of measurable process outputs that can be used to quantify the problem as it currently exists, and the setting of

specific goals and schedules for the resolution of the problem. The third step requires a listing of all potential sources of variation (e.g., by a cause and effect diagram) and information on what is known about each source (known or unknown cause of variation, possible magnitude of its effect, responsibility for its control, etc.). This information can be used to assign priorities to the sources of variation. The objective here is to select a few of the highest priority sources for the initial study.

To illustrate our discussion, consider the following simplified process of a machining operation as displayed in Figure 1. The variation in shaft diameters from a final machining operation has caused assembly problems downstream and a continuous improvement effort is undertaken to reduce this variation. The machine carrying out this operation has two grinding heads performing the same operation and receives parts from a rough grinding machine. Both heads have the same lubricant and power supply and the same operator is in charge throughout a shift. There are two shifts per day and approximately four rod lots are used each shift. The first three steps of the problem solving process have produced the following:

- selected problem: reduce diameter variation
- process description: as displayed in Figure 1
- measured output: diameter
- input identification and prioritization:

operator	grinding head
rod lot	tool change
lubricant temperature	power supply
diameter after rough grind	

The next step is to plan the data collection scheme for investigating the effects of the prioritized sources. We assume in what follows that the chosen sampling scheme does not involve

a deliberate process change. That is, it is an observational study because no intervention is made in the process. Besides observational studies, designed experiments is the other major approach for data collection. Reasons for preferring observational studies over experiments are that there is no interference with production, no risk of producing poor quality output by deliberately changing the process, low cost, etc. Note that observational studies have long been used by the social and health sciences for investigating human populations (Cochran (1983)), where it is often impossible to conduct a designed experiment. To the best of our knowledge, this paper is the first to propose using observational studies in the industrial setting for continuously improving processes.

## **Sampling Designs for Different Sources of Variation**

We assume that the outputs of the process to be measured have been identified along with the potential sources of variation or inputs; that is, the first three steps in the problem solving process described in the previous section have been done. (We will use interchangeably the terms input and source of variation.) Using the Pareto principle for sources of variation, the objective of the study is to identify those inputs whose changes transmit large changes to the outputs. There are two obstacles, however. First, while a large number of inputs have been listed, only a few can be simultaneously dealt with in an investigation. This is why the inputs must first be prioritized and then only the highest of these be considered. Second, the “current noise” in the process may cause so much variation in the output that the transmitted change induced by an input is masked. As we will describe, sampling schemes, purposely designed, can help to alleviate this difficulty.

Since different views of the process require different sampling schemes, their construction begins by selecting a key input to be investigated; it is a bonus if other inputs can also be investigated with the same scheme. It is important to recognize that, to a large degree, the nature of the input determines the sampling plan. With this in mind, we classify the inputs into three broad types:



- Stratification (*ST*): Examples are machines, lines, pallets, shifts, operators, fixtures, etc. Each stratum usually has a fixed effect on the output and can be sampled from at different time periods. The strata are usually determined qualitatively. Also, it can easily be determined in which stratum the process is operating.
- Scheduled Changes (*SC*): Examples are batches, lots, tool changes, etc. These inputs are characterized by a predictable length of time between input changes (i.e., a scheduled change) and have a variable effect on the output from one input change to another. The input changes are usually defined qualitatively.
- Unscheduled Changes (*UC*): Examples are lubrication temperature, line pressure, viscosity, etc. Characteristics of these inputs are that the length of time between input changes is unpredictable, the input changes are defined quantitatively and the inputs can be measured on-line.

A source of variation cannot be identified unless it is traceable. It must be possible to identify the operators, machines, lots, etc. which produce each sampled output. For unscheduled inputs, the input value corresponding directly to each sampled output must be determined. For example, if humidity is not monitored, it can never be identified as a source of variation.

Regardless of the input's nature, there are two basic principles that should be followed when designing a sampling scheme.

**I. Keep as many other inputs as possible constant while the key input under study changes.**

**II. Replicate I for as many changes in the other inputs as possible.**

Principle I is equivalent to blocking in designed experiments which reduces the noise component of the output so that the magnitude of the key input effect is more apparent. Principle II is the same as replication in designed experiments and provides the opportunity to assess the key input's effect on the output under varying production conditions. By replication, the consistency of the effect can be examined. Thus, observational studies can incorporate two of the guiding principles used in designing experiments.

Next, sampling plans employing Principles I and II will be described for investigating the prioritized inputs of the grinding operation. The classification of these sources is:

operator	( <i>ST</i> )	grinding head	( <i>ST</i> )
rod lot	( <i>SC</i> )	tool change	( <i>SC</i> )
lubricant temperature	( <i>UC</i> )	power supply	( <i>UC</i> )
diameter after rough grind	( <i>UC</i> )		

### **Sampling Designs for Stratification Inputs**

First, we consider a sampling design for grinding head and then operator.

#### **Example 1: Grinding Head**

Many inputs affect both heads simultaneously. The sampling plan consists of selecting two consecutive rods from the rough grind operation and finishing them simultaneously on the two heads (Principle I). Repeat this procedure to obtain outputs from pairs representative of production (Principle II); i.e., over as many different level combinations of the other inputs as possible.

The selection of pairs of rods as described in the sampling scheme holds all other inputs constant while the two rods are being finished. It is assumed that before the rough grinding

operation the two consecutive rods had a similar past history.

**Example 2: Operator**

Very few inputs have a common effect on both operators' output. Among the prioritized inputs, only grinding head has a constant effect on output while operators change. The sampling plan consists of the following. Each time during the investigation that the process is operating within an operator, collect a sample representative of the output from each head (Principle I). Repeat this for each operator over a range of levels for the other inputs (Principle II).

The sampling scheme in Example 1 is a suitable design if the pairs of rods are selected to give a sample representative of the output during the period of time an operator is on duty. Moreover, this sampling scheme provides the opportunity to investigate the effect of both *ST* inputs. It is easier to detect head-to-head differences than operator-to-operator differences, however, since more other inputs are held constant as heads change than as operators change. Note that the operator effect cannot be distinguished from that of another input which changes from shift-to-shift. For example, the effect of an operator for a previous step in the process is confounded with that of the final machining operator. If the same operators always work together, there is no way of separating their effects; this can only be accomplished by an intervention.

**Sampling Designs For Scheduled Change Inputs**

A useful aid in designing a sampling scheme for key scheduled change inputs is to make a time-line of when the these inputs change. A time-line for the machining operation is given below.

operator 1				operator 2				
lot 1	lot 2	lot 3	lot 4	lot 5	lot 6	lot 7	lot 8	...
tool 1	tool 2	tool 3	tool 4	tool 5	tool 6			

Note that lot and tool changes are nested within operator changes in this scenario. Also, both heads operate throughout all changes of the prioritized inputs.

### **Example 3: Lot**

It is assumed that within tool change there is a tool wear effect; i.e., the tool wears and it affects the output. If the output is randomly sampled during the period when a lot is being machined then the lot effect will be confounded with the tool wear effect. A more effective sampling strategy is as follows. For the two consecutive lots within a tool change, observe the output for the last  $n$  pairs (say 2 to 5, where the pairs refer to the two rods being simultaneously ground on the two heads) of the first lot and the first  $n$  pairs of the second lot (Principle I). Repeat this over a number of shifts (Principle II).

It is crucial that lot number, tool change number and shift be traceable. It is also assumed that the last  $n$  rods and the first  $n$  rods are representative of their respective lots. Note that this sampling plan is not appropriate for assessing an operator effect.

### **Example 4: Tool Change**

Here, we consider the difference between the rod diameters produced just preceding and following a tool change so that the tool change effect is a composite of several effects including tool wear and tool differences. The sampling plan consists of observing the output for the last  $n$  pairs before a tool change and the first  $n$  pairs after the tool change. (Principle I). Repeat this over a number of shifts (Principle II).

Note that this sampling plan is not appropriate for assessing either an operator or lot effect.

## **Sampling Designs For Unscheduled Change Inputs**

Recall that unscheduled change inputs must be traceable. It may require hard work and some ingenuity to devise a tracking system. For example, if the diameter of the rough ground shaft is the key input in the machining operation, a tracing method is needed to

relate the incoming diameter to the diameter after final grind. A sampling strategy for this unscheduled change input is considered next.

#### **Example 5: Diameter After Rough Grind**

The sampling plan consists of sampling within head and operator, measuring the diameter before and after the final grind (Principle I). Repeat over all heads and operators for a number of lots. Collect sufficient data to prepare scatter plots within each stratum of the  $ST$  inputs. (Principle II).

A suitable sampling scheme to assess the effects of the incoming diameter is the one suggested for the stratified inputs. This same scheme is also suitable for assessing the effects of lubrication temperature and power supply.

#### **Some Comments**

The machining operation example demonstrates how the two basic design principles can be used to develop sampling schemes. Note that the same sampling scheme could be used to assess several of the inputs and might have been easily overlooked unless careful planning had taken place. Other inputs may require different sampling plans, however.

## **Analysis**

In the proposed approach, the sampling scheme is designed to assess the effect of a specific input. As a consequence, the analysis will depend on the sampling scheme and can be done off-line. Note that even when several inputs require the same sampling scheme, their respective analysis will be different. The analysis can usually be done using the simple graphical methods that form part of Ishikawa's (1982) seven basic tools: trend charts, control charts and scatter plots.

Recall that Principle I for designing sampling schemes is that data should be collected in blocks if possible. Within each block, output measurements for each key input value are

collected while as many of the remaining inputs as possible are kept constant. This principle suggests that the analysis for assessing whether the *ST* and *SC* inputs affect the output average be a comparison of the output averages for the different key input values within each block. By collecting data for a number of blocks (Principle II), a consistent pattern in these averages across the blocks can be looked for; such a pattern suggests that the input has an effect.

In the machining example, when grinding head (Example 1) is the key input, differences between the heads can be plotted. A pattern of points consistently above or below zero in the plot would suggest a head effect. For example, see Figure 2 which plots the difference between head 1 and head 2 and illustrates the situation where head 2 produces larger diameter rods. The effect could then be confirmed computationally with a paired t-test by using the variation in the differences of individual pairs of head data as a measure of noise.

When operator (Example 2) is the key input, averages for each operator within a shift can be plotted separately for each head. The averages for both heads may also be combined and plotted for each operator across shifts. One operator's points falling consistently above another operator's suggests an operator effect. For the chart with heads plotted separately, any apparent patterns could be confirmed computationally in this case with an unpaired t-test using the variation of the output within head and operator as a measure of noise. Note that this requires that several replicates be taken within a single shift, i.e., the output needs to be sampled more than once within a single shift. For situations when there are more than two strata (e.g., operators), the strata means can be compared graphically using Ott's (1975) analysis of means or computationally by an appropriate analysis of variance.

Separate plots of operator averages for each head have an additional advantage. They provide an assessment of the consistency of an operator effect for each head, i.e., an interaction effect between head and operator may be suggested. Also note that while the effects of *ST* inputs on the output average has been the focus, their effects on the output variation can be assessed with R charts. For example in the machining example, separate R charts of

within operator ranges for each head could be used to assess the effect of head and operator on the output variation.

For *SC* inputs, such as lots in the machining example, the effect on the output varies from one change to another so that there is no point in looking for trends. The pertinent question is whether, within a sampling block for lots, the between-lot variation exceeds that expected from the within-lot variation. This question can be addressed by constructing appropriate  $\bar{X}$  and R charts. Recall that for each head, the last  $n$  rods before the lot change and the first  $n$  rods after the change were sampled. The R chart is constructed by first computing the ranges for these “subgroups” of size  $n$ , plotting them and calculating the appropriate upper control limit. See Figure 3 for an example which illustrates a process where the within lot variation is consistent. To assess the between-lot variation (i.e., a lot effect), construct an “ $\bar{X}$  chart” by plotting the difference of averages between the two lots in each sampling block. The center line is zero and the control limits account for the difference of averages (rather than a single average) being plotted (i.e.,  $\pm\sqrt{(2)}A_2\bar{R}$ , where  $A_2$  is the appropriate constant (Grant and Leavenworth (1988)) and  $\bar{R}$  is the average range from the R chart). Many out-of-control points such as those displayed in Figure 3 indicate a lot effect.

The effects of *UC* inputs can be assessed by scatterplots of output versus input values. Principles I and II imply that there should be enough data available for separate scatterplots for each stratum of an important *SC* input. Consequently, interactions between *ST* and *UC* inputs can be studied; plotting the data for both heads on the same graph using different plotting symbols to distinguish between the heads may reveal different relationships between output and temperature (i.e., different intercepts or slopes).

Note that in the series of analyses described above, a subsequent analysis of an input may be influenced by the analysis of a previously studied input. For example, if an analysis of an *ST* input suggests that it has an effect, subsequent analyses of the remaining inputs might be done separately for each strata of that *ST* input.

We have presented some general approaches for analyzing data from the various input

types and related them to the machining example. The particular methods are necessarily problem specific. We have emphasized a graphical approach which is simple to do, easy to interpret and therefore, a useful first pass at analyzing the data. We have also mentioned more sophisticated analyses that can detect smaller input effects and can be pursued if necessary.

## Summary

In this paper, we have concentrated on designing sampling schemes to identify specific sources of variation; different “views of the process” are needed to assess the effects of different sources. The analysis of the data is done off-line and is largely based on the seven basic tools. This approach provides a powerful strategy for making the “diagnostic” portion of Juran’s journey. While control charts are often suggested as a tool for this journey, they only take one view of the process. Moreover, we believe that control charting is more appropriate for providing a process record and for maintaining stability by signaling the need for adjustments.

In contrast with designed experiments, this non-interventionist approach costs less, is easier to implement and does not interrupt production. Note, however, that confounding of effects may present interpretation problems, the process cannot be changed to assess an effect, the natural range of variation may not be large enough to detect an effect and the duration of the study may be longer than that for a designed experiment. The first problem may be alleviated by rerouting the flow of some of the samples, i.e., an intervention, although perhaps a minor one. For example, suppose that the rough grind operation in the machining example also has two heads and rods finished in head 1 of the final grind normally come from rough grind head 1. Then some rods from the other rough grind head will need to be rerouted to the final grind head 1 in order to decouple the potential head effects of the two operations.

Continuous improvement requires an iterative approach. Regardless of what strategy



is used, improvement will not take place unless the “diagnostic journey” is followed by a “remedial journey.” As sources of variation are eliminated or their effects reduced, new sources will become important. In order to assess these new sources, their own special sampling scheme will be needed.

We have emphasized the necessity of upfront planning. In conclusion, we believe that the lack of such planning and purposeful sampling has resulted in the ineffective use of charting and explains a large part of the frustration that many have experienced using it.

### **Acknowledgments**

We thank Jeff Wu for helpful comments on an earlier version. M. Hamada was supported in part by research grants from General Motors of Canada Limited, the Manufacturing Research Corporation of Ontario, and the Natural Sciences and Engineering Research Council of Canada.

### **References**

- Cochran, W. G. (1983). *Planning and Analysis of Observational Studies*. John Wiley and Sons, New York, NY.
- Deming, W. E. (1986). *Out of the Crisis*. Massachusetts Institute of Technology Center for Advanced Engineering Study, Cambridge, MA.
- Grant, E. L. and Leavenworth, R. S. (1988). *Statistical Quality Control*. McGraw-Hill Book Company, New York, NY.
- Ishikawa, K. (1982). *Guide to Quality Control*. Asian Productivity Organization, Tokyo, Japan.
- Juran, J. M. and Gryna, Jr., F. M. (1980). *Quality Planning and Analysis*. McGraw-Hill Book Company, New York, NY.

MacKay, R. J. (1988). "Continuous Improvement With Control Charts". Research Report RR-88-04, University of Waterloo Institute for Improvement in Quality and Productivity, Waterloo, Ontario.

Ott, E. R. (1975). *Process Quality Control*. McGraw-Hill Book Company, New York, NY.

Pyzdek, T. (1990). "There's No Such Thing as a Common Cause". *ASQC Quality Congress Transactions*. American Society for Quality Control, pp. 102-108.

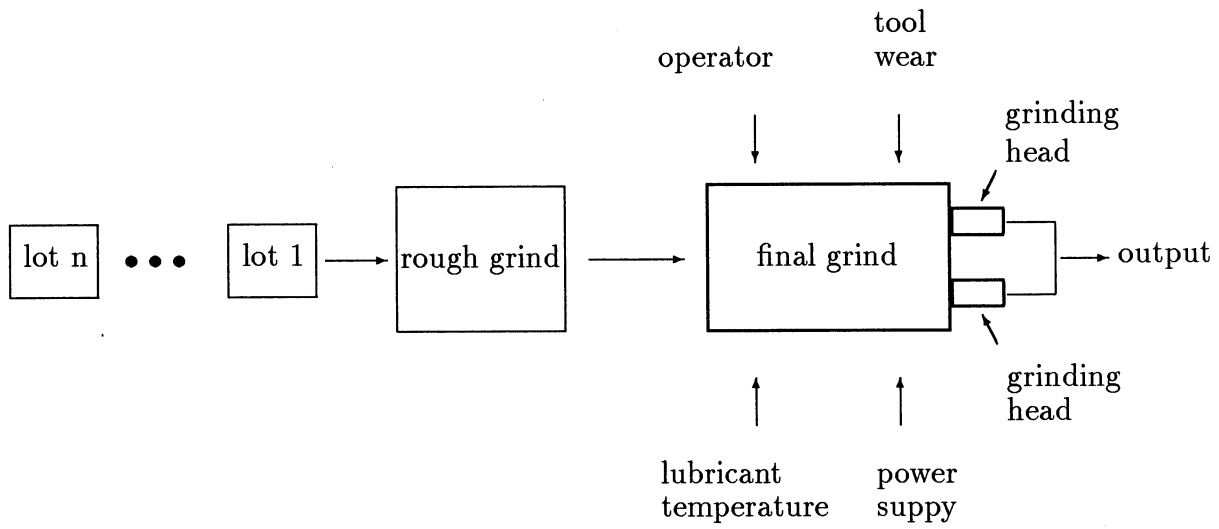
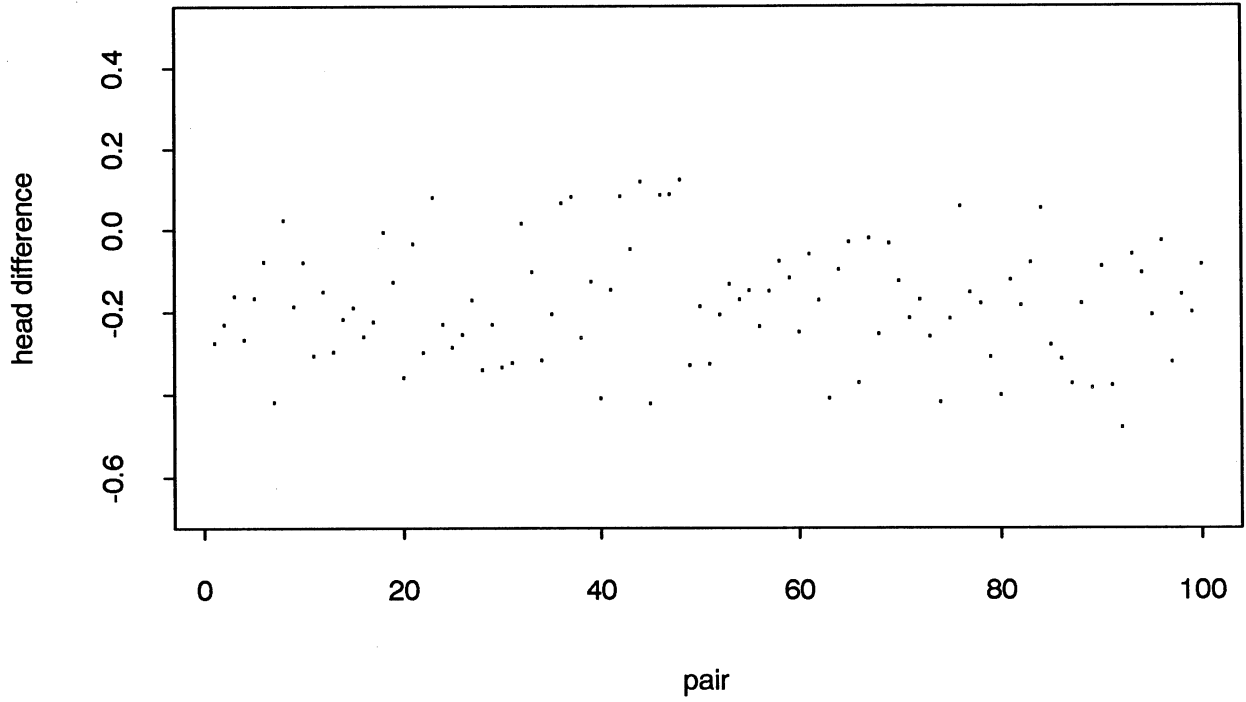


Figure 1: Process Flow Diagram

Figure 2: Head Difference for 5 Days  
10 random pairs per shift



**Figure 3: Lot Ranges and Mean Differences Between Lots for 10 Days**  
2 pairs per sample

