STATISTICS IN ADVANCED MANUFACTURING

Vijay Nair[1], Mark Hansen and Jan Shi[2]

University of Michigan and Bell Laboratories

Statistical concepts and methods have played a critical role in speeding the pace of industrial development over the last century. In return, industrial applications have provided statisticians with incredible opportunities for methodological research. The richness and variety of these applications have had a major influence on the development of statistics as a discipline: consider, for example, the extensive research in statistical process control (SPC) and change-point detection, dating back to the pioneering work of Shewhart in the 1920's, developments in automatic process control, design of experiments, sequential analysis, reliability, and so on. Recent efforts by manufacturers to adopt sustained quality and productivity improvement programs have generated a renewed interest in and appreciation for statistics in industry. In fact, fundamental statistical concepts such as understanding and managing variation form the backbone of popular quality management paradigms and practices.

Many of the traditional SPC concepts and techniques grew in response to the manufacturing environments prevalent several decades ago. Current advanced manufacturing and high-technology industries, however, operate under a much more complex and diverse set of conditions. These changes have important implications for research directions in industrial statistics, not only in terms of identifying new problems and developing new methods, but also in re-evaluating the paradigms that inspired earlier approaches. In this vignette, we use applications from automotive and semiconductor manufacturing to illustrate various issues and to discuss future research needs and directions. The discussion is limited to a few selected topics and is inevitably slanted towards our own experiences.

THE ENVIRONMENT OF ADVANCED MANUFACTURING

Pressures from the competitive marketplace are forcing manufacturers to continuously reduce product development cycle times. In parallel, the underlying technology of the products and processes are becoming increasingly complex in order to keep up with consumer demands. Thus, manufacturers frequently move from design to full-scale production well before the technology and the fabrication processes are completely understood.

[1]V. N. Nair's research was supported by NSF-DMS grant 9803281.

[2]Jan Shi's research was supported by NSF CAREER grant: DMI 9624402.

Consider the manufacture of semiconductor devices. The critical dimensions of integrated circuits (IC's or chips) are very fine: newly developed IC's have features as small as $0.16\mu m$, and any company hoping to remain competitive will have plans to reduce these sizes to as little as $0.01\mu m$ in the next $5-10$ years. Given the scale of these devices, IC's are fabricated in a "cleanroom" through a process involving hundreds of separate steps and lasting several weeks. The various steps are rarely "stable," as is commonly assumed in the SPC paradigm, and they frequently interact in unexpected ways. Such complexity and instability are typical in advanced manufacturing applications.

VOLUME AND COMPLEXITY OF THE DATA

Massive amounts of process and product quality data are now collected routinely, made possible by advances in both computing and data acquisition technologies. Much of these data have special structures; images, functional data, marked point processes, and high-dimensional time series are all common.

In IC fabrication, several hundred chips are fabricated simultaneously on a *wafer*, and the wafers are themselves processed in groups called *lots*. Large manufacturing plants can start thousands of wafers a week. A wide range of measurements are made on each wafer, including particle data, in-line electrical measurements, and final probe test data even before the wafers are shipped to be packaged. The final probe test alone generates a vector of $15-20$ dimensional measurement for each chip. In all, one collects as much as 1.5Gb of data per week in a typical fab. It is well known that problems in different manufacturing steps will leave telltale spatial signatures (which can vary within and across lots), so one has to treat the wafer map data as inherently spatial objects.

In the automotive industry, reducing autobody "dimensional" variation is one of the major quality challenges. Autobody assembly involves several hundred parts and more than 100 assembly stations. With the implementation of in-line optical coordinate measurement machines (OCMM), tremendous amounts of dimensional data are now routinely collected. The OCMM measures 100 to 150 points on each major assembly with a 100% inspection rate. These data exhibit both spatial and temporal dependence.

Both the volume and the complexity of the data dictate the need for fast and flexible methods of analysis as well as appropriate environments for computing and visualization.

PROCESS MONITORING FOR DATA WITH COMPLEX STRUCTURE

While there has been a tremendous amount of research in process monitoring, much of it has

focused on new and yet more powerful tests for detecting changes in process means and variances. The real need in advanced manufacturing applications, however, is in dealing with processes where the observations have complex structure. Due to the lack of appropriate statistical methods and software tools for analyzing such data, practitioners typically force the problem into a more traditional framework, often resorting to simple overall summary measures that ignore the structure in the data. In so doing, we lose valuable opportunities for process improvement.

Figure 1 is a tonnage signal collected from a sheet metal stamping process. The signal corresponds to a complete forming cycle and measures the total forming force from all the tonnage sensors mounted on the stamping press. This is a typical example of the kind of *functional* data that are now being collected and used to monitor and diagnose problems in manufacturing processes. A traditional approach that treats the data as a vector of multidimensional observations and applies standard multivariate SPC techniques has been shown to be quite inadequate in this application (see Jin and Shi, 1999). We will return to this application later in the vignette.

Figure 2 describes why spatial patterns are important in IC data. The rightmost wafer in this figure is a display of binary (pass/fail) probe test results collected at the end of the fabrication process. This map can be viewed as the superposition of the two wafer maps on the left: a cluster of defective chips representing a special or assignable cause, with a pattern that helps to identify the responsible machine or area; and (essentially) random defects resulting from the overall cleanliness of the fabrication line. A traditional SPC approach would simply summarize the observed test results with a single measure, wafer yield. Such an analysis clearly misses critical spatial information about yield loss. Hansen, Nair and Friedman (1997) describe methods for monitoring binary, spatial processes to detect the presence of spatial clustering. As the null hypothesis of complete spatial randomness is too simplistic, they use a a Markov random field to characterize null situations with mild spatial clustering. This spatial process monitoring effectively complements the information from control charts that only track yield. However, as shown in Hansen, Nair and Friedman (1999), this spatial monitoring tool is only the *first step* in fully exploiting the spatial character of these data.

BEYOND PROCESS MONITORING: THE REAL OPPORTUNITIES FOR STATISTICS

Intelligent Failure Diagnostics:

The primary emphasis in SPC has been on monitoring and change point detection. The development of failure diagnostics and root cause determination have typically been considered to be the

domain of experts with subject matter knowledge. In advanced manufacturing applications where the processes are not well understood, even the subject matter experts are increasingly relying on in-process and product quality data for diagnosing process problems. From a statistical point of view, there is more information in the data when a process goes out of control than when it is in control. This is where the real opportunities are and where we can make important contributions. Unfortunately, statisticians have been slow in recognizing these opportunities, and much of the exciting work is being done by nonstatisticians in the engineering community.

To provide a concrete example of the role that statistics can play in this area, consider again the IC application. The monitoring scheme for spatial data discussed earlier will trigger an alarm when there is significant spatial clustering as in Figure 2, but it will not provide any information about the *nature* of the clusters. As discussed in Hansen, Nair and Friedman (1999), spatial patterns provide tremendous information about potential process problems. For example, if the defects concentrate in the center of the wafer, there is likely a problem controlling the thickness of a chemical "resist" deposited on the surface of the wafer prior to lithography. Hansen, Nair and Friedman (1999) describe statistical techniques for extracting *spatial signatures* of defect patterns and using them to immediately identify one or more likely root causes. These methods have been successfully deployed within Lucent Microelectronics. Since our original involvement in this area, many new techniques have appeared, and SEMATECH recently sponsored an entire conference on spatial statistics and pattern recognition in IC fabrication (see `www.sematech.org/public/resources/stats/Symposium/1999/index.htm`).

Combining Detection and Diagnosis:

In manufacturing environments with high-volume, there is a need to diagnose process problems as they occur in *real time*. The usual two-stage approach of detection followed by diagnosis, typically done off-line, is not adequate in this setting. Ideally, we should integrate statistical methods with underlying engineering knowledge about potential faults and failure diagnostics to develop a combined approach to detection and diagnosis. It is reasonable to view this as a classification problem where the different classes represent different possible faults or states of the "system" (including the null state). An initial specification of these states can be obtained during the design and development stage, and they can be constantly updated as we collect and analyze on-line process and product quality data.

Jin and Shi (1999) provide a good example of this in the context of the tonnage signal data

for stamping processes (Figure 1). Their methodology segments the tonnage signal according to the different stages of the forming cycle and exploits information about the potential faults and how they will manifest themselves on the the tonnage signal. For example, a flat peak is the result of a loose tie rod while an oscillating peak indicates a worn bearing (see Figure 1). Jin and Shi (1999) describe a wavelet-based statistical analysis for feature extraction and use these features to do process monitoring and fault diagnosis.

Ceglarek and Shi (1996) and Apley and Shi (1998) describe another application from the automotive industry involving "fixture" failure diagnostics. Design and maintenance of the fixturing process is an important problem as the "dimensional" variation in the autobody panels depends critically on the quality of fixturing process. Ceglarek and Shi (1996) describe how the engineering knowledge about the fixture geometry and tooling design layout can be used to develop a model that relates variations in the sensor measurements to different fixture faults. Apley and Shi (1998) and Dong (1999) show that incorporating this information provides tremendous advantage over traditional SPC methods, not only in being able to detect the faults quickly but also in being able to diagnose the problems in real time.

In on-going research on the IC application, we are studying methods for developing a library of spatial templates that characterize different process problems and for classifying observed wafer maps with spatial clustering into one of the templates.

DATABASES, COMPUTING AND VISUALIZATION

As mentioned above, advances in sensing and data acquisition technologies have made it possible to routinely collect massive amounts of data about manufacturing processes. Statisticians must be involved in all aspects of this data collection process, ensuring that the right kinds of data are being collected and stored, helping design appropriate measurement systems (choices of sensors, their location and number, etc.), assuring data quality, and so on. We have seen two trends in data acquisition and storage over the last few years.

First, the various streams of data currently pool into different data bases. Process control engineers have one source for machine-level routing and maintenance information, while yield enhancement engineers pull post-production tests results from another. In recent years, there has been a move to centralize data collection, management and access so that factory-wide information can soon be readily applied to process improvement efforts. In addition to creating methods that make use of these new data sources (an incredible challenge in itself), statisticians have an

important role to play in helping design and implement effective data warehousing solutions.

The second trend we have observed relates to the *information content* being stored in factory data bases. As the both the complexity and volume of data increase, space and computing considerations dictate that some form of reduction take place prior to storage in a data base. At a practical level, this can represent a gain in efficiency because the reduced or *compressed* data might be more readily amenable to SPC methods. See, for example, Jin and Shi (1999) for the use of statistical techniques for feature-based data compression in stamping processes. In general, we expect that statisticians will be called upon to design specialized compression techniques for storing only the most relevant information. This poses an incredible challenge in that unlike traditional notions of "sufficiency" for a parametric model, departures from standard operating conditions can be quite complex and often difficult to anticipate.

Finally a comment on the role of statistical computing and graphics in advanced manufacturing. Naturally, the practical success of any industrial application depends largely on acceptance of new statistical methods by engineers in the factory. In our IC work, visualization was critical, as was the development of a convenient computing environment in which to express quantitative ideas about in-line and post-production fabrication data. This led to the creation of a software platform, called S-wafers, which makes use of the object-based facilities of the S language. As a vehicle for technology transfer, S-wafers supports a range of tasks from generating automated reports (disseminated via the Lucent intranet in the form of html documents and Java applets) to interactive, specialized analyses on one or more lots. Engineers familiar with S can immediately augment their routine data analysis (based on a summary like yield) with spatial information. We believe that our experience is not unique and that industrial statisticians who take seriously the challenges of technology transfer are regularly called upon to make use of and to create new and novel tools for computing and visualization that can be transferred to the engineers responsible for manufacturing.

SUMMARY

We close by mentioning two other challenges that face industrial statisticians as we look ahead to the next century. First, a major consequence of the drive to reduce product development cycle times is that manufacturers are moving away from physical experimentation and testing to computer modeling and CAD/CAM tools. This presents a whole array of research opportunities for statisticians, ranging from model validation and verification to efficient design and analysis of very high-dimensional computer (or virtual) experiments.

Second, statisticians have too often worked in isolation and developed fragmented approaches that ignore important information and interactions present in sequential, multi-stage manufacturing processes. The complexities of advanced manufacturing environments dictate that we must work closely with engineers and take a system's approach to process improvement. Ceglarek, Shi, and Wu (1994), for example, describe a knowledge-based system for "dimensional" control that has been very effective in improving the quality of new autobody assembly processes. The methodology captures information about the multi-level assembly process and factory layout through a hierarchical structural model. Combining statistical analysis of the in-process data with information about the assembly architecture and sequence allows for the root causes of process variability to be diagnosed quickly and efficiently.

As all these examples suggest, statistics has an extremely important role to play in industry as we move into the data-rich, information era of the 21st century. These are indeed exciting times for our profession with a wide array of interesting research opportunities.

REFERENCES:

Apley, D. and Shi, J. (1998), " Diagnosis of Multiple Fixture Faults in Panel Assembly," *ASME Transactions, Journal of Manufacturing Science and Engineering,* Vol.120, 793-801.

Ceglarek, D. and Shi, J. (1996), "Fixture Failure Diagnostics for Autobody Assembly Using Pattern Recognition," *Journal of Engineering for Industry*, Vol. 118, 55-66.

Ceglarek, D., Shi, J. and Wu, S. M. (1994), "A Knowledge-Based Diagnostic Approach for the Launch of the Auto-Body Assembly Process," *ASME Transactions, Journal of Engineering for Industry,* Vol. 116, 491-499.

Dong, X. (1999), "Contributions to Statistical Process Control and Robust Parameter Design," Ph. D. Dissertation, University of Michigan.

Jin, J, and Shi, J., (1999) "Automatic Feature Extraction of Waveform Signals for In-Process Diagnostic Performance Improvement," *Technometrics* (to appear).

Hansen, M.H., Nair V.N., and Friedman, D.J., (1997) "Monitoring Wafer Map Data in Integrated Circuit Fabrication for Spatially Cluestered Defects," *Technometrics*, Vol.39, No.3, 241-253.

Hansen, M., Nair, V, and Friedman, D., (1999), "Process Improvement Through the Analysis of Spatially Clustered Defects on Wafer Maps," JASA (submitted).

Figure 1: A tonnage signal collected from a sheet metal stamping process. One complete forming cycle is displayed. The response is the total forming force measured by all the tonnage sensors mounted on the stamping press.

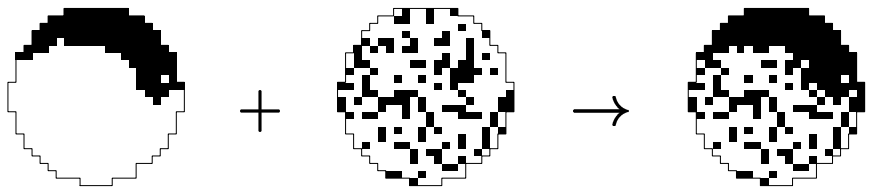


Figure 2: *A graphical model for overall wafer yield. The defective chips in the middle wafer occur essentially at random, while those on the left are "clustered," reflecting a process problem. The final probe map is a superposition of these two processes.*