Improving manufacturing quality through planned experiments: statistical methodology

Markus Abt and **Friedrich Pukelsheim,** Augsburg

Summary. Planned experimentation is a powerful tool to improve quality of industrial manufacturing processes, in an environment placing prime concern on top quality. The goals are to manufacture products that function best possible, to achieve a minimum variability of the production process, and to secure product robustness against noise in the customer's environment. Statistically planned experiments are a tool to achieve these goals. They consist of a few experimental runs to obtain data on the product characteristics. Each run corresponds to a planned choice of levels of production factors and noise factors. Depending on the experimental data, the production factors are classified into signal factors that in the first place affect the process mean, control factors that predominantly determine process variability, and nuisance factors that can be used for cost reduction. Noise factors are an additional category, they mirror the customer's environment. Any such cycle of manufacturing improvement should terminate with a confirmation experiment.

AMS Subject Classification: 62K, 62P.

Keywords: Dispersion analysis; Linear statistical model; Location analysis; Method of least squares; Power transformations; Response surface designs.

1 Introduction

Quality improvement of manufacturing processes by planned experiments has a long history. Box and Draper (1969) describe this approach under the heading of evolutionary operation; Daniel (1976) is another established source. Box and Draper (1987) and Taguchi (1987) are more recent references. In this paper we review the features that make this approach so successful. The two companion papers Weihs, Berres and Grize (1994), and Abt, Mayer and Pukelsheim (1995) illustrate how the method works in practice.

In Sect. 2 we argue that Quality Engineering is part of a more embracing concept, total quality management. Section 3 on Planned Experiments makes a point that great progress can be made with such well-known statistical methods as the analysis of variance. The design objectives (Sect. 4) for improving manufacturing quality serve a dual purpose, to bring the process mean on target, and to minimize process variability.

The analysis of the experimental data leads to a factor classification (Sect. 5). According to the statistical analysis of the experimental data, the production factors that

are under supervision of the process engineers are classified into signal factors, control factors, and nuisance factors. A planned experiment can also involve a fourth category, noise factors, that influence product performance in the customer's environment. A confirmation experiment (Sect. 6) terminates the proposed cycle of quality improvement, to make sure whether the statistician's recommendations satisfy the needs of the process engineers.

The subject has received renewed attention through the work of Taguchi (1987), although many of the issues raised can be identified in the publications of Box (1985), or Deming (1986). A bibliography of books, special journal issues, symposia, articles and case studies of recent years is included in Pukelsheim (1988, 1991). Textbook treatments are Logothetis and Wynn (1989), John (1990), Grove and Davis (1992). Specific aspects are discussed in Pukelsheim (1986), León, Shoemaker and Kacker (1987), Box (1988), Ghosh (1990), Vining and Myers (1990), Welch, Yu, Kang and Sacks (1990), Shoemaker, Tsui and Wu (1991), Eibl, Keß and Pukelsheim (1992), Hamada and Wu (1992), Nair (1992). Banks (1993) offers a current discussion of the subject.

2 Quality engineering

Statistical tools to improve manufacturing quality may be subsumed under the broader heading of *quality engineering*. On a management level, product quality has to rank top. Producing at low costs, meeting the production plan, or improving worker productivity are issues of lower importance.

Quality engineering is obsessed with data. In order to pin down the strong and weak points of the production process, data are an absolute necessity. Without data, commitment evaporates into verbal eloquence, discussions degenerate into departmental fights, and instructions fail to be operational.

The classical methods of quality assurance emphasize acceptance sampling and sampling schemes, and process control and control charts. This is a passive approach, preventing the worst by weeding out bad products. In contrast, quality engineering is a continuous activity, to build quality into the product from the very beginning.

Quality engineering embraces the whole life time of a product, starting as early as in the design stage (off-line), and continuing on into the manufacturing stage (on-line). It extends through the time the product has to function in the customer's environment, and it terminates only when the product is scrapped. Thus quality engineering goes far beyond the use of planned experiments.

Improvement by experiment and analysis, prediction and confirmation is just one aspect of quality engineering. The statistical methods that we describe below are easy

to use, yet sophisticated enough to pick up typical process behavior. They accomodate the standard engineering situation that many factors enter into a manufacturing process. By viewing these factors in their entirety, a vast improvement is achieved compared to naïve experiments that vary just one factor at a time.

In a nutshell our message is two-fold: Planned experiments form a valuable methodology, easy to use, yet strong in content. On the other hand, they are just one tool in the strategic kit called quality engineering.

3 Planned experiments

Planned experimentation is one of the most established ways of scientific analysis. Draper and Pukelsheim (1996) provide a state of the art survey of the design of experiments, including the statistical prerequisites, and covering applications in industry, agriculture, and biometry.

Planned experiments for quality improvement of manufacturing processes often come under the heading of response surface methodology. They consist of the totality of statistical techniques that apply when numerical data from an industrial process is represented by a "best fit" mathematical surface.

Suppose, for example, that we are studying the effect of changing the production factors temperature (t_1), pressure (t_2), and viscosity (t_3) on the yield (y) of a chemical process. Ideally the production factors determine a *true response* $\eta(t_1, t_2, t_3)$. Practically, this true value is distorted by an additive measuring error (e). Hence the model assumption for the observed yield is

$$y = \eta(t_1, t_2, t_3) + e. \tag{1}$$

The error e is assumed to come from a specific error distribution, typically a normal distribution with mean 0 and variance $\sigma^2 > 0$.

The true response function $\eta(\mathbf{t})$ of the production conditions $\mathbf{t} = (t_1, t_2, t_3)'$ is usually unknown, a prime denoting transposition. Hence the statistician or engineer attempts to model (approximate) the true relationship by some member from a function class $g(\mathbf{t}, \theta)$ depending on an unknown parameter vector θ:

$$\eta(\mathbf{t}) = g(\mathbf{t}, \theta). \tag{2}$$

When the graduating function g depends linearly on θ, that is $g(\mathbf{t}, \theta) = f(\mathbf{t})'\theta$, we obtain the usual linear statistical model:

$$y = f(\mathbf{t})'\theta + e. \tag{3}$$

The essential ingredient to completely specify a model of type (3) is the regression function f.

For example, a three-factor second-degree *model* approximates the unknown true response surface using a quadratic polynomial in three variables,

$$\begin{aligned} y = \theta_0 + t_1\theta_1 + t_2\theta_2 + t_3\theta_3 + t_1^2\theta_{1,1} + t_1t_2\theta_{1,2} + t_1t_3\theta_{1,3} \\ + t_2^2\theta_{2,2} + t_2t_3\theta_{2,3} + t_3^2\theta_{3,3} + e. \end{aligned} \tag{4}$$

The regression function $f(t_1, t_2, t_3)$ consists of the powers $1, t_1, t_2, t_3, t_1^2, t_1t_2, t_1t_3, t_2^2$, t_2t_3, t_3^2. Accordingly the vector θ consists of the 10 unknown real parameters $\theta_0, \theta_1, \theta_2$, $\theta_3, \theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{2,2}, \theta_{2,3}, \theta_{3,3}$. If $\theta_{i,j} = 0$ for $i \leq j$ is assumed in (4), a *first-degree model* is specified that fits a planar surface of the 3 factors t_1, t_2, t_3.

The standard technique in these situations is to conjecture a model, often based on prior experience, to choose an experimental design, to run the experiment and obtain the data, and to analyse the data and check whether the statistical model fits the experimental evidence reasonably well.

An experimental design consists of runs $u = 1, \ldots, n$, where run u is determined by a vector $\mathbf{t}_u$ of production factor settings. It is assumed, at least to practical approximation, that the production factors can be set without error. For each run u, a yield y_u is observed. The data $\mathbf{t}_1, \ldots, \mathbf{t}_n$ and $y_1, \ldots, y_n$ are then used to fit a model of type (3), by estimating numerical values $\widehat{\theta}$ for the parameter vector θ, often with the method of least squares. The result is the estimated response surface

$$\widehat{y} = f(\mathbf{t})'\widehat{\theta}, \tag{5}$$

and the corresponding *fitted values* $\widehat{y}_u = f(\mathbf{t}_u)'\widehat{\theta}$. Even if the model assumption (3) is justified, the fits $\widehat{y}_u$ are affected by the random errors e_u. Of course, model (3) itself may be inappropriate, and then the estimated response surface $\widehat{y}$ provides a more or less biased view of the true response surface that underlies the production process.

Usually the simpler, first-degree model is an appropriate choice to start with. A variety of tests on the estimated response surface $\widehat{y}$ can be made. Most of these diagnostic procedures rely on the unexplained portions of the observed yield, $y_u - \widehat{y}_u$, called the *residuals*. If the planar model does not fit well, a second-degree surface would be tried.

More details can be found in the survey paper of Draper and Pukelsheim (1996) that was mentioned above. Textbook treatments that offer a general introduction into the statistical design of experiments are Box, Hunter and Hunter (1978), Draper and Smith (1981), Box and Draper (1987), or Khuri and Cornell (1987).

The practical choice of an experimental design is guided by many (and partly conflicting) goals, see the discussion in Box and Draper (1987, Chapter 15). One way of selecting a design is to concentrate on its symmetry properties, of which rotatability is the most important one for the case of response surface designs.

Another approach is to choose a design according to whether it maximizes or minimizes a specific criterion function, within a class of competing designs. Bandemer (1977) and Bandemer and Näther (1980) are encyclopedic volumes, with extensive tabulations of many important design families. Pukelsheim (1993) presents a general optimality theory of experimental designs in linear models. The results on optimal, theoretical designs can then be used to identify efficient, practical designs.

4 Design objectives

The long-standing approach that has stood the test of time is to model the process yield, y, through a linear statistical model (3). This aims at the mean process output. The production factors that prove to be significant are then used to bring the mean process yield on target, maximize it, or minimize it, as desired. This goal is achieved using the statistical methods that are provided by the classical analysis of variance, and response surface methodology, see e.g. the textbooks of Box and Draper (1987), Daniel (1976), Khuri and Cornell (1987), Logothetis and Wynn (1989).

However, Taguchi (1987) draws attention to a dual application of the same statistical methods, to also model process variability. Other than in the agricultural and biometrical sciences, in industrial experimentation it is often feasible to obtain independent replications of each run. Let n_u be the replication number, that is the number of observed yields given the production factor settings $\mathbf{t}_u$. We denote the corresponding observed process yields by $y_{u,j}$, for $j = 1, \ldots, n_u$. That is $y_{1,1}, \ldots, y_{1,n_1}$ are the n_1 observations under experimental conditions $\mathbf{t}_1$, etc.

The estimated variance under experimental conditions $\mathbf{t}_u$ then is

$$s_u^2 = \frac{1}{n_u - 1} \sum_{j=1}^{n_u} (y_{u,j} - \overline{y}_u)^2, \tag{6}$$

with average yield $\overline{y}_u = \sum_{j=1}^{n_u} y_{u,j}/n_u$. Ideally the quantities s_u^2 have a scaled χ^2 distribution. Hence the logarithm is applied as a variance stabilizing transformation,

$$z_u = \log s_u^2 \qquad \text{for } u = 1, \ldots, \ell, \tag{7}$$

provided there are ℓ different sets of settings $\mathbf{t}_1, \ldots, \mathbf{t}_\ell$ for the production factors. Taguchi (1987) fits a linear model to $z_1, \ldots, z_\ell$, in order to study how process variability depends on the production factors.

Thus the experimental data are evaluated through a dual analysis, in the sense of Vining and Myers (1990). For the process mean, a location model is built to fit the observed yields $y_{1,1}, \ldots, y_{\ell,n_\ell}$. For the process variability, the same statistical methodology leads to a dispersion model for the log sample variances $z_1, \ldots, z_\ell$. These dual analyses have proved extremely successful in practice.

In theory, the distributional assumptions underlying a linear model can be fully satisfied either for the random variables $y_{1,1}, \ldots, y_{\ell,n_\ell}$, or else for the random variables $z_1, \ldots, z_\ell$. However, linear statistical models have transpired to be useful also when the underlying assumptions are satisfied only to some degree of approximation. Applying the logarithm to the sample variances as in (7) does indeed serve as a normalizing transformation. In a more general context, this is the guiding principal for the Box–Cox power transformations of the data when building a linear model, as proposed by Box and Cox (1964). Other transformations can also be contemplated, as León, Shoemaker and Kacker (1987).

5 Factor classification

As a result of dually analyzing process mean and process variability, the production factors are classified into three categories. The first category are the signal factors, consisting of those factors that are significant in the model for the process mean and nonsignificant in the model for the process variability. Hence the signal factors may be used to bring the process mean on target, maximize it, or minimize it, as desired, without having a statistically recognizable effect on process variability.

The second category, control factors, comprises those production factors that are significant in the process variability model. The model is used to predict settings for the control factors that minimize the production variance. In other words, the process is "kept under control" not just by monitoring control charts, but actively identifying the set of control factors among the set of all production factors, and by implementing the optimum levels for the control factors so as to minimize process variability.

Control factors should be clearly designated as such, so that they are not accidentally changed by process operators. Also, when the process variability is observed to increase, it is a good policy to first check whether the control factors have moved away from their optimum levels.

The classification in signal factors and control factors is data-dependent. It can thus not be excluded that occasionally the observed data lead to empty or otherwise unreasonable categories. According to our experiences it would seem that the classification goes through successfully in most practical cases.

The remaining production factors, that appear neither among the signal factors nor among the control factors, are called nuisance factors. They are nonsignificant for

the process mean, nor for the process variability. If no other goal intervenes, nuisance factors are set to minimize production costs.

In addition to production factors, a planned experiment may also incorporate noise factors and study their effect on the product. Noise factors are factors in the user environment that are vital for the product to function properly. For example, humidity, surrounding temperature, or exposure to daylight would not generally occur as genuine production factors for a manufacturing process. But they may well appear as noise factors once the product is installed at the user's site.

With noise factors included, the analysis of the experimental data hopefully points to settings for the production factors that perform well over the range of values that the noise factors may attain. As a result, an experiment including process factors as well as noise factors leads to a manufacturing process for a product that is robust once the buyer starts to use it.

In other words, planned noise involves a source of variation that in a way is artificial from the viewpoint of the proper manufacturing process. However, from the point of view of product quality, this variability is deliberately introduced in order to mimic environmental noise. In this way the experiment anticipates variability of the product performance, in a laboratory setting, long before use of the product has actually started.

6 Confirmation experiment

The confirmation experiment is the formal end of a cycle of statistical quality improvement as described above. It implements the predicted optimum levels for the production factors, and "proves" them to be superior to the old settings. The confirmation experiment reassures all those involved in the project that their efforts were well invested and start paying back.

References

Abt, M., Mayer, R., and Pukelsheim, F. (1995). "Improving manufacturing quality through planned experiments: pressure governor case study." *Surveys on Mathematics for Industry* **5** 35–47.

Bandemer, H. (Editor) (1977). *Theorie und Anwendung der optimalen Versuchsplanung I. Handbuch zur Theorie.* Akademie-Verlag, Berlin.

Bandemer, H. and Näther, W. (1980). *Theorie und Anwendung der optimalen Versuchsplanung II. Handbuch zur Anwendung.* Akademie-Verlag, Berlin.

Banks, D. (1993). "Is industrial statistics out of control?" *Statistical Science* **8** 356–377 [Discussion, 378–409].

Box, G.E.P. (1985). *The Collected Works of George E.P. Box.* (G.C. Tiao, C.W.J. Granger, I. Guttman, B.H. Margolin, R.D. Snee, S.M. Stigler, Editors). Wadsworth, Belmont, CA.

Box, G.E.P. (1988). "Signal-to-noise ratios, performance criteria, and transformations." *Technometrics* **30** 1–17 [Discussion, 19–40].

Box, G.E.P. and Cox, D.R. (1964). "An analysis of transformations." *Journal of the Royal Statistical Society Series B* **26** 211–243 [Discussion, 244–252].

Box, G.E.P. and Draper, N.R. (1969). *Evolutionary Operation.* Wiley, New York.

Box, G.E.P. and Draper, N.R. (1987). *Empirical Model-Building and Response Surfaces.* Wiley, New York.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978). *Statistics for Experimenters.* Wiley, New York.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation.* Wiley, New York.

Deming, W.E. (1986). *Out of the Crisis.* Cambridge University Press, Cambridge, UK.

Draper, N.R. and Pukelsheim, F. (1996). "Design of experiments: a survey." *Statistical Papers* **37** 1–32.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis.* Wiley, New York.

Eibl, S., Keß, U. and Pukelsheim, F. (1992). "Achieving a target value for a manufacturing process: a case study." *Journal of Quality Technology* **24** 22–26.

Ghosh, S. (1990). *Statistical Design and Analysis of Industrial Experiments.* Dekker, New York.

Grove, D. and Davis, T.P. (1992). *Engineering, Quality, and Experimental Design.* Longman, Harlow, UK.

Hamada, M. and Wu, C.F.J. (1992). "Analysis of designed experiments with complex aliasing." *Journal of Quality Technology* **24** 130–137.

John, P.W.M. (1990). *Statistical Methods in Engineering and Quality Assurance.* Wiley, New York.

Khuri, A.I. and Cornell, J.A. (1987). *Response Surfaces.* Dekker, New York.

León,R.V., Shoemaker, A.C. and Kacker, R.N. (1987). "Performance measures independent of adjustment." *Technometrics* **29** 253–265 [Discussion, 266–285].

Logothetis, N. and Wynn, H.P. (1989). *Quality through Design.* Clarendon Press, Oxford.

Nair, V.J. (1992). "Taguchi's parameter design: a panel discussion." *Technometrics* **34** 127–161.

Pukelsheim, F. (1986). "Experimental design and off-line quality control." In *Proceedings of the Holzhau 1986 Seventh International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis. Seminarbericht Nr. 84* (O. Bunke, Editor). Sektion Mathematik der Humboldt Universität, Berlin, 134–139.

Pukelsheim, F. (1988). "Analysis of variability by analysis of variance." In *Optimal Design and Analysis of Experiments* (Y. Dodge, V. Fedorov and H.P. Wynn, Editors), North-Holland, Amsterdam 1988, 281–292.

Pukelsheim, F. (1991). "Neuere statistische Methoden für die Versuchsplanung bei industriellen Fertigungsprozessen." *Zeitschrift für Angewandte Mathematik und Mechanik* **71** T568–T573.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, New York.

Shoemaker, A.N., Tsui, K.-L. and Wu, C.F.J. (1991). "Economical experimentation methods for robust design." *Technometrics* **33** 415–427.

Taguchi, G. (1987). *System of Experimental Design.* Kraus, New York.

Vining, G.G. and Myers, R.H. (1990). "Combining Taguchi and response surface philosophies: a dual response approach." *Journal of Quality Technology* **22** 38–45.

Weihs, C., Berres, M. and Grize, Y.-L. (1994). "Statistical design of experiments in industrial practice." *Surveys on Mathematics for Industry* **5** 49-73.

Welch, W.J., Yu, T.-K., Kang, S.M. and Sacks, J. (1990). "Computer experiments for quality control by parameter design." *Journal of Quality Technology* **22** 15–22.

Authors' address: Dr. M. Abt and Dr. F. Pukelsheim, Institut für Mathematik, Universität Augsburg, D-86135 Augsburg, Federal Republic of Germany