Surveys on Mathematics for Industry **5** (1995) 35–47

Improving manufacturing quality through planned experiments: pressure governor case study

Markus Abt, Robert Mayer, and **Friedrich Pukelsheim,** Augsburg

Summary. The paper reports on a planned experiment to improve pressure governor quality in a large scale manufacturing process. The presentation provides an outline to conduct similar industrial experiments for quality improvement of manufacturing processes.

AMS Subject Classification: 62K, 62P.

Keywords: Confirmation experiment; Control factor; Dispersion analysis; Factor classification; Industrial experimentation; Location analysis; Location–dispersion plot; Main effects model; Noise factor; Normal plots; Orthogonal array; Production factors; Screening experiment; Signal factor.

1 Introduction

Planned experiments are an important tool for statisticians and engineers to improve product quality. General aspects, and applications in the chemical and pharmaceutical industries are presented in the companion papers Abt and Pukelsheim (1995) and Weihs, Berres and Grize (1995). In the present paper we report on a fairly large experiment that we carried out to improve an already established production line.

We have set out the material so as to serve as an example for similar situations. On the other hand we supply sufficient detail so that an interested reader can repeat our analysis, or carry out a different one.

2 The Product

A pressure governor is a technical device to take some fluid or gas as input, at varying pressure, and output the same substance, at constant pressure. It consists of a bottom part, the pot, a top part, the cap, and an in-between rubber membrane. The incoming and outgoing fluid circulates through the lower chamber, consisting of the pot and sealed off by the membrane. The upper chamber in the cap contains the mechanical parts to equalize pressure variation.

The vital quality characteristic of a pressure governor is burst pressure. It is measured by increasing the pressure in the lower chamber until the membrane fails. This destructive test is carried out on a few items per batch. If the results are satisfactory then the batch is shipped to the customer.

Pressure governor production is a complex process, with many input factors, involving quite a few departments of the company. The manufacturing process is well understood from the engineering point of view, and a voluminous documentation was available. However, increasing experience with the process did not result in improved quality. Rather, burst pressure seemed to decrease over time, if only slightly so.

The in-house quality assurance group had already carried out many one-factor-at-a-time experiments, *within* the single departments that contributed to the manufacturing process. From this experience it became evident that the process needed to be investigated as a whole, and it was decided to run a larger experiment *across all* departments involved.

The goal was two-fold, to find settings for the production factors that would lead to a high burst pressure, while at the same time guaranteeing a low process variability.

3 The Factors

In order to find out which production factors to include in the experiment, extensive interviews and discussions with the personnel from all departments were necessary. Naturally, each department favored the factors that they contribute to the process. On the other hand local expertise may become a hindrance for a global viewpoint, in particular when a project involves many different groups like in the present case. The final agreement was to investigate 22 factors that were considered to be of potential significance to the manufacturing process. They split up into five major groups.

The first group consisted of factors affecting the surface finish of pot and cap, chroming type, chrome layer thickness, finish roughness, brightness additive, galvanization process.

The next two factor groups are concerned with the mechanicals of fitting the cap on the pot. In the final production step, the membrane is laid on the pot, and the cap is laid on the membrane. Then the larger pot rim is bent around the smaller cap rim. The second group, the cap factors, are metal thickness, rim angle, and number of grooves on the side of the cap rim where the membrane goes. The third group had five pot factors: pot type, rim shape, number of grooves on the side of the rim that faces the membrane, height and soldering time of the in- and output valves.

The fourth group is concerned with the rubber membrane: supplier, hardness, thickness, and sliding additive. The membranes are cut from large rubber sheets of which the top and bottom sides have a different texture. The factor roughness described whether the rough side of the membrane faced the cap or the pot.

The fifth group of factors was associated with the final assembly, to bend the larger pot rim over the smaller cap rim, with the membrane in-between: amount of pressure, pressure speed, press type, and dampening coefficient.

Once the 22 factors had been singled out, we decided on the distinct levels with which they would appear in the experiment. It so happened that half of them were run at two levels, and the other half at three levels. For the quantitative factors the levels are interpreted as low and high, or low, middle, and high. In Table 1, they are coded as $-, +$, or $-, 0, +$. For the qualitative factors different levels are just different names, such as one supplier versus another supplier. For easy perception, however, we use the same codings $-, +$, or $-, 0, +$.

Most factors were proper production factors. That is, process engineers were eligible to change these factors at their discretion. However, some factors do not qualify as production factors, but rather appear as noise factors, such as pot type, or rubber supplier. Different pot types are manufactured to the client's order. In the case of rubber suppliers, business relations were equally excellent and there was no reason to forego any one of them.

In terms of the experiment, the goal was to optimize the level settings for the production factors, across the whole range of levels for the noise factors. However, as engineers wanted to be able to recognize potential differences between various pot types, rubber suppliers, or so on, we decided to model the problem *without* discriminating between production factors and noise factors. In this way we also obtained estimates for the effects of the different levels of the noise factors.

4 The Model

A particular experimental design receives its merits only against the intended statistical analysis. It is therefore imperative to outline the statistical model that is to be employed, or at least to delineate the class of models that might be entertained.

As it was generally agreed that out of the 22 factors some would be more important than others, our predominant task was to screen the factors according to their significance for the production process. For this reason we decided to use a *main-effects model*, with linear effects for the two-level factors $t_1, \ldots, t_{11}$, and linear and quadratic effects for the three-level factors $t_{12}, \ldots, t_{22}$. That is, our model was of the form

$$y = \theta_0 + t_1\theta_1 + \cdots + t_{22}\theta_{22} + t_{12}^2\theta_{12,12} + \cdots + t_{22}^2\theta_{22,22} + e. \tag{1}$$

This model leaves 35 coefficients to be estimated, the constant θ_0, the 22 first-degree effects $\theta_1, \ldots, \theta_{22}$, the 11 second-degree effects $\theta_{12,12}, \ldots, \theta_{22,22}$, and the variance σ^2 of the random error term e.

Inclusion of *all* interaction effects $\theta_{i,j}$ for the interactions $t_i t_j$ of the factors t_i and t_j would have led to an excessive increase in the number of model parameters. Inclusion of only *some* interactions in the model might have been an option, but we felt we did not have sufficient information to do so.

5 The Design

Model (1) has 34 unknown parameters for the mean, whence we needed at least this many different runs in the experiment. To satisfy this need we chose for the experimental design the orthogonal array OA $(36, 2^{11} \times 3^{12}, 2)$, as shown in Table 1. The design is due to Taguchi (1960), and also given on page 415 of Logothetis and Wynn (1989). It consists of 36 rows, called *runs*, and is able to support 11 factors at two levels and 12 factors at three levels. The design is of strength 2, that is, in any two out of the 23 columns all possible pairs occur equally often. The statistical meaning is that the design is of resolution class III. This means that just main effects are unbiasedly estimable, while already two-factor interactions are not estimable, but confounded with some main effects.

Each row of the design determines a way of setting the levels of the 22 factors $t_1, \ldots, t_{22}$. Two-level factors are coded by $-$ or $+$: low or high. For three-level factors the levels are indicated by $-, 0$, or $+$: low, medium, or high. These levels correspond to actual settings that the process operators could implement within the available region of operability, at no extra cost. For the quantitative three-level factors, the levels corresponded to equi-spaced settings. It so happened that run 25 was almost identical to the *standard operating conditions* that were in use at the time when our cooperation began.

The presentation in Table 1 emphasizes the combinatorial structure of the design. For the practical implementation, a time sequence was chosen that made the level changes from one run to the next technically easy. The ensuing ordering is indicated by the superscript numbers in the first column. In other words, there is a severe lack of randomization in the way the experiment was carried out, in terms of runs as well as in terms of replications.

6 The Data

The number of items tested within each run was set at 10. One reason was that during the interview stage we failed to obtain unanimous information on the process variability. Secondly, we resorted to the fact that the pressure governors are produced in high numbers, and that an individual device is relatively cheap. Thirdly the destructive test arrangement consisted of a rack that was able to test 10 items simultaneously.

Run	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}	t_{11}	t_{12}	t_{13}	t_{14}	t_{15}	t_{16}	t_{17}	t_{18}	t_{19}	t_{20}	t_{21}	t_{22}	t_{23}
1^{26}	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
2^{24}	−	−	−	−	−	−	−	−	−	−	−	0	0	0	0	0	0	0	0	0	0	0	0
3^{22}	−	−	−	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+
4^{27}	−	−	−	−	−	+	+	+	+	+	+	−	−	−	0	0	0	0	+	+	+	+	−
5^{25}	−	−	−	−	−	+	+	+	+	+	+	0	0	0	+	+	+	+	−	−	−	−	0
6^{23}	−	−	−	−	−	+	+	+	+	+	+	+	+	+	−	−	−	−	0	0	0	0	+
7^{13}	−	−	+	+	+	−	−	−	+	+	+	−	0	+	−	0	+	+	−	0	0	+	−
8^{14}	−	−	+	+	+	−	−	−	+	+	+	0	+	−	0	+	−	−	0	+	+	−	0
9^{15}	−	−	+	+	+	−	−	−	+	+	+	+	−	0	+	−	0	0	+	−	−	0	+
10^{19}	−	+	−	+	+	−	+	+	−	−	+	−	+	0	−	+	0	+	0	−	+	0	−
11^{20}	−	+	−	+	+	−	+	+	−	−	+	0	−	+	0	−	+	−	+	0	−	+	0
12^{21}	−	+	−	+	+	−	+	+	−	−	+	+	0	−	+	0	−	0	−	+	0	−	+
13^{18}	−	+	+	−	+	+	−	+	−	+	−	−	+	−	+	0	−	+	+	0	−	0	0
14^{17}	−	+	+	−	+	+	−	+	−	+	−	0	−	0	−	+	0	−	−	+	0	+	+
15^{16}	−	+	+	−	+	+	−	+	−	+	−	+	0	+	0	−	+	0	0	−	+	−	−
16^{12}	−	+	+	+	−	+	+	−	+	−	−	−	+	0	−	−	+	0	+	+	0	−	0
17^{11}	−	+	+	+	−	+	+	−	+	−	−	0	−	+	0	0	−	+	−	−	+	0	+
18^{10}	−	+	+	+	−	+	+	−	+	−	−	+	0	−	+	+	0	−	0	0	−	+	−
19^{4}	+	−	+	+	−	−	+	+	−	+	−	−	−	+	+	+	−	0	0	−	0	+	0
20^{5}	+	−	+	+	−	−	+	+	−	+	−	0	0	−	−	−	0	+	+	0	+	−	+
21^{6}	+	−	+	+	−	−	+	+	−	+	−	+	+	0	0	0	+	−	−	+	−	0	−
22^{1}	+	−	+	−	+	+	+	−	−	−	+	−	0	+	+	−	0	−	−	+	+	0	0
23^{2}	+	−	+	−	+	+	+	−	−	−	+	0	+	−	−	0	+	0	0	−	−	+	+
24^{3}	+	−	+	−	+	+	+	−	−	−	+	+	−	0	0	+	−	+	+	0	0	−	−
25^{34}	+	−	−	+	+	+	−	+	+	−	−	−	0	−	0	+	+	−	+	−	0	0	+
26^{35}	+	−	−	+	+	+	−	+	+	−	−	0	+	0	+	−	−	0	−	0	+	+	−
27^{36}	+	−	−	+	+	+	−	+	+	−	−	+	−	+	−	0	0	+	0	+	−	−	0
28^{7}	+	+	+	−	−	−	−	+	+	−	+	−	0	0	0	−	−	+	0	+	−	+	+
29^{8}	+	+	+	−	−	−	−	+	+	−	+	0	+	+	+	0	0	−	+	−	0	−	−
30^{9}	+	+	+	−	−	−	−	+	+	−	+	+	−	−	−	+	+	0	−	0	+	0	0
31^{28}	+	+	−	+	−	+	−	−	−	+	+	−	+	+	0	+	0	0	−	0	−	−	+
32^{29}	+	+	−	+	−	+	−	−	−	+	+	0	−	−	+	−	+	+	0	+	0	0	−
33^{30}	+	+	−	+	−	+	−	−	−	+	+	+	0	0	−	0	−	−	+	−	+	+	0
34^{33}	+	+	−	−	+	−	+	−	+	+	−	−	−	0	+	0	+	−	0	0	+	−	+
35^{32}	+	+	−	−	+	−	+	−	+	+	−	0	0	+	−	+	−	0	+	+	−	0	−
36^{31}	+	+	−	−	+	−	+	−	+	+	−	+	+	−	0	−	0	+	−	−	0	+	0

Table 1: Orthogonal array $\mathrm{OA}(36, 2^{11} \times 3^{12}, 2)$. The runs $r = 1, \ldots, 36$ determine the factor levels low ($-$), medium (0), or high ($+$). The plan supports 11 factors $t_1, \ldots, t_{11}$ at two levels, and 12 factors $t_{12}, \ldots, t_{23}$ at three levels (of which t_{23} was not used). Strength 2 means that in any two columns all possible pairs occur equally often.

With 10 replications for each of the runs $r = 1, \ldots, 36$, we obtained a total of $n = 360$ observed responses $y_{r,j}$, in units of the scale marks on the measuring device. For the purpose of communication we refer to them in *bar*. The data are given in Table 2.

A convenient way of a first, informal analysis of the data is to draw boxplots for each experimental run, see Figure 1. We *did not* find sufficient evidence to identify any one of the observed responses as an outlier and exclude it from further analysis.

The boxplots visibly show that the mean location and the spread of the 10 repli-

Run	$y_{r,1}$	$y_{r,2}$	$y_{r,3}$	$y_{r,4}$	$y_{r,5}$	$y_{r,6}$	$y_{r,7}$	$y_{r,8}$	$y_{r,9}$	$y_{r,10}$	$\bar{y}_r \pm s_r$	
1^{26}	124	152	146	132	122	122	154	144	146	114	135.6 ± 14.4	
2^{24}	151	136	128	116	133	123	134	133	121	131	130.6 ± 9.7	
3^{22}	142	144	147	150	130	134	141	152	132	147	141.9 ± 7.7	
4^{27}	131	122	131	133	131	139	130	130	124	133	130.4 ± 4.7	
5^{25}	134	132	132	131	140	136	144	136	134	132	135.1 ± 4.1	
6^{23}	134	130	108	122	123	132	130	122	112	105	121.8 ± 10.3	
7^{13}	195	205	202	192	200	220	204	202	195	202	201.7 ± 7.7	fairly good run
8^{14}	174	153	176	134	190	164	184	162	180	176	169.3 ± 16.5	
9^{15}	162	169	186	194	192	175	164	210	152	202	180.6 ± 19.1	
10^{19}	174	166	184	184	176	170	156	171	181	205	176.7 ± 13.1	
11^{20}	91	112	121	90	131	111	150	118	106	101	113.1 ± 18.2	smallest $\bar{y}_r$
12^{21}	202	216	233	244	206	216	196	164	212	208	209.7 ± 21.5	largest $\bar{y}_r$, largest s_r
13^{18}	180	174	164	184	180	166	182	170	172	170	174.2 ± 7.0	
14^{17}	172	172	182	172	172	182	176	172	178	174	175.2 ± 4.1	
15^{16}	137	123	130	134	116	132	130	125	131	144	130.2 ± 7.7	
16^{12}	146	124	134	150	142	114	148	112	162	161	139.3 ± 17.9	
17^{11}	142	111	136	140	155	140	142	124	110	142	134.2 ± 14.6	
18^{10}	127	144	152	130	125	155	142	154	141	143	141.3 ± 10.9	
19^{4}	182	174	164	184	182	174	171	182	175	172	176.0 ± 6.4	
20^{5}	142	137	142	138	158	122	150	146	143	135	141.3 ± 9.6	
21^{6}	182	156	152	172	162	170	170	122	160	162	160.8 ± 16.2	
22^{1}	172	176	178	182	182	176	172	182	193	170	178.3 ± 6.8	
23^{2}	185	204	170	201	170	200	195	208	174	190	189.7 ± 14.3	
24^{3}	111	112	120	122	114	114	114	125	134	121	118.7 ± 7.2	
25^{34}	148	142	111	132	131	133	146	132	145	121	134.1 ± 11.7	standard conditions
26^{35}	185	203	191	207	194	205	213	214	204	181	199.7 ± 11.4	
27^{36}	142	132	126	132	140	132	140	138	134	132	134.8 ± 5.0	
28^{7}	154	132	105	130	142	148	134	136	134	127	134.2 ± 13.2	
29^{8}	166	150	172	174	182	172	184	176	182	174	173.2 ± 9.9	
30^{9}	142	122	116	138	151	148	156	145	122	121	136.1 ± 14.6	
31^{28}	164	162	154	140	166	158	162	172	168	152	159.8 ± 9.3	
32^{29}	122	121	114	128	122	115	121	121	120	121	120.5 ± 3.9	smallest s_r
33^{30}	152	143	142	152	147	148	132	147	142	144	144.9 ± 5.8	
34^{33}	151	174	152	149	160	156	154	132	185	162	157.5 ± 14.4	
35^{32}	166	182	178	180	154	185	182	172	174	158	173.1 ± 10.6	
36^{31}	178	182	160	162	174	181	174	177	176	161	172.5 ± 8.4	

Table 2: Burst pressure data. The responses $y_{r,j}$ are observed per run $r = 1, \ldots, 36$, each with 10 replications. The actual time sequence in the experiment is indicated by the superscript numbers in the first column. Also given are the average $\bar{y}_r$ and the sample standard deviation s_r, within run r.

---

cations varied considerably from run to run. Run *7* was fairly good, both in terms of average response as well as small spread. Run *11* featured the smallest average response, run *12* the largest one together with the largest spread. Run *25*, standard operating conditions, was not among the best runs. Nor was it among the worst runs, of course. Run *32* had the smallest spread, similar to runs *4*, *5*, and *14*, but neither achieved an acceptable process average.

From this preliminary data analysis, one could settle with the factor settings of run *7*. The observed average burst pressure of 202 bar, with a sample standard deviation of 8 bar, is a clear improvement over the standard operating conditions of

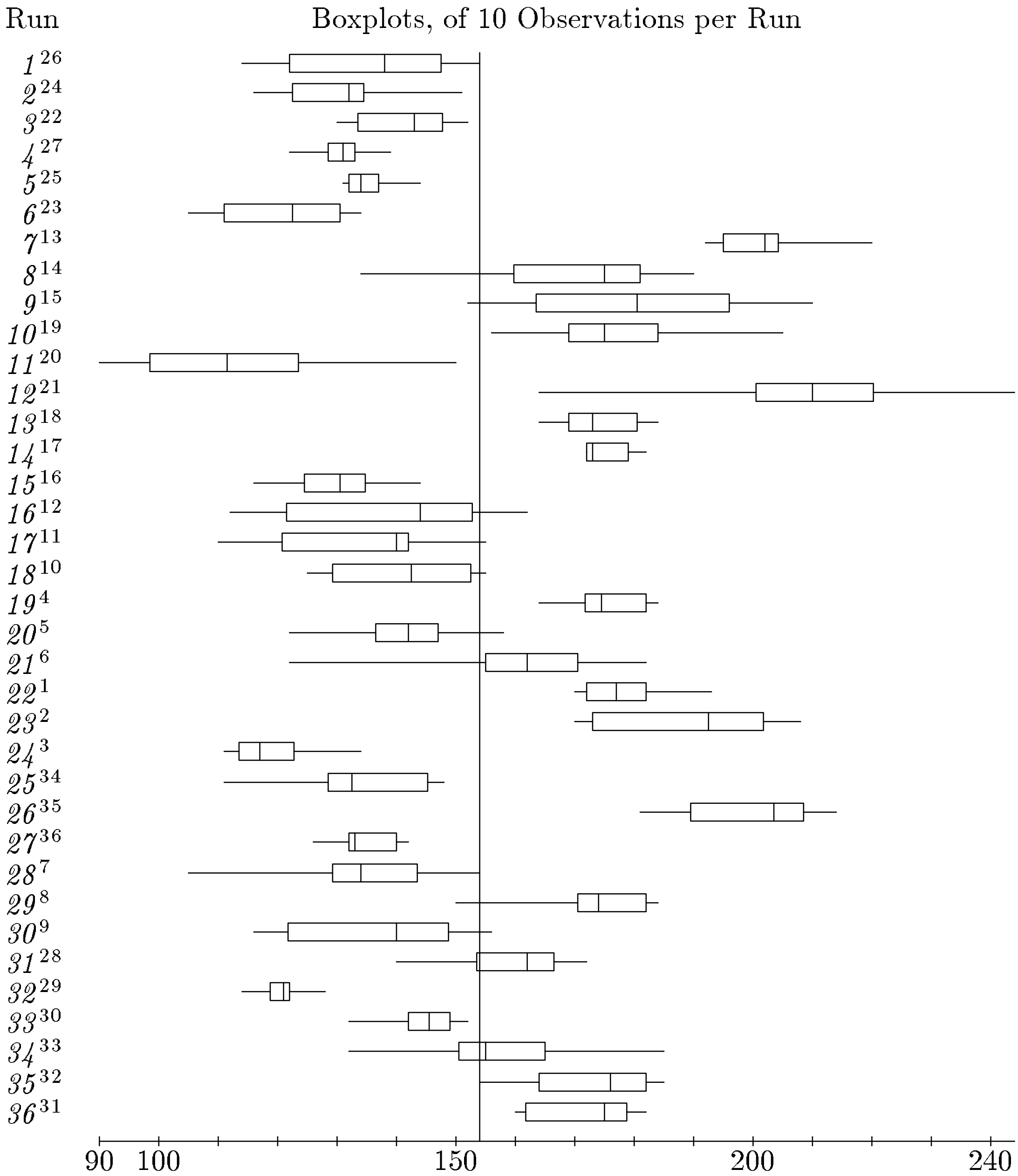

Fig. 1: Per run boxplots. For each run, the left whisker starts with the smallest response and ends at the first quartile. The box in the middle is subdivided by the median observation. The right whisker starts at the third quartile and ends with the maximum. The vertical line is the overall average of 154 bar.

---

run *25*, with an average of 134 ± 12 bar. However, this would base our inference on just 10 observations, and disregard the remaining 350 values. Fitting to the full data set a model such as (1) exploits all 360 observed responses, to obtain a much more complete picture of the response surface.

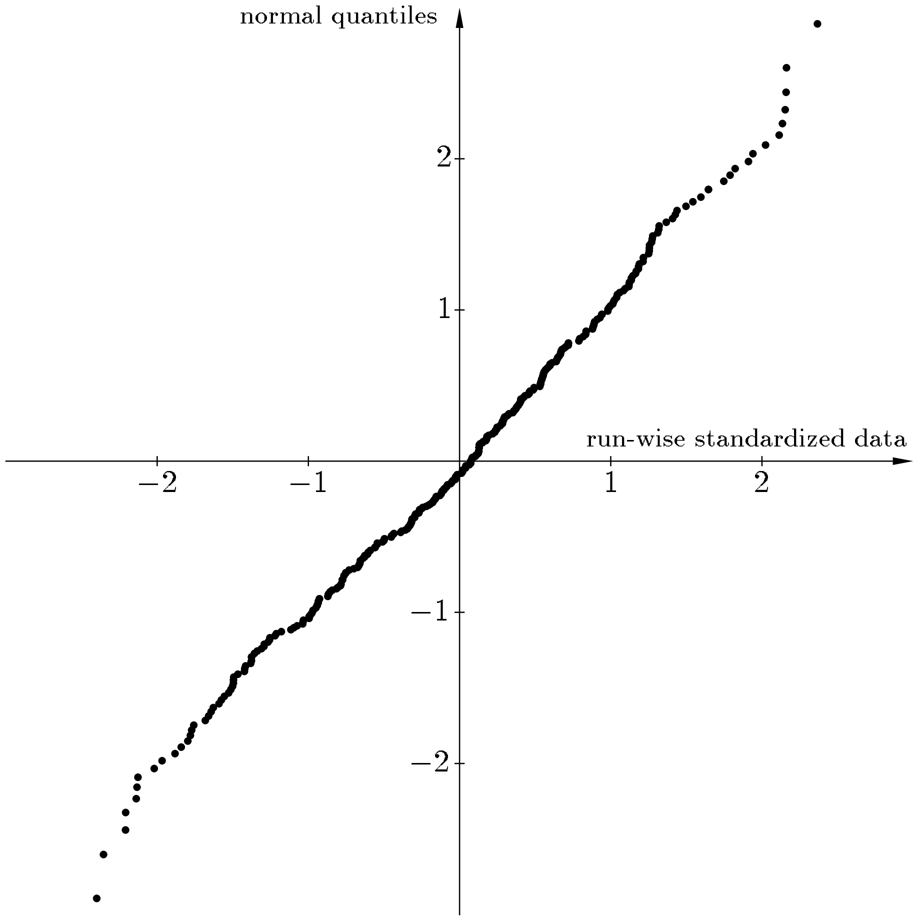

Fig. 2: Normal plot of standardized burst pressure data. Heteroscedasticity across runs calls for a standardization of the observations $y_{r,j}$. The 360 new values $x_{r,j} = (y_{r,j} - \bar{y}_r)/s_r$ are ordered and plotted against the $(k-0.5)/360$ standard normal quantiles. The high correlation of 0.9973 conforms with a normality assumption.

7 Data Transformations

A routine analysis of model (1) requires that the observations are normally distributed. If this applies to the numbers $y_{1,1}, \ldots, y_{36,10}$ that we observed, it *will not* be true for such transformations as $\sqrt{y_{r,j}}$ or $(y_{r,j})^2$ that might have been reported just as well. Or the other way round, if $\sqrt{y_{r,j}}$ or $(y_{r,j})^2$ happen to be normally distributed, then $y_{r,j}$ cannot possibly also be normal.

Accordingly we ran our data through a loose test whether the normality assumption on $y_{r,j}$ is justified. Within each run $r = 1, \ldots, 36$, we calculated the mean response and the sample standard deviation,

$$\overline{y}_r = \frac{1}{10}\sum_{j=1}^{10} y_{r,j}, \qquad s_r^2 = \frac{1}{9}\sum_{j=1}^{10}(y_{r,j} - \overline{y}_r)^2.$$

The standardized versions $x_{r,j} = (y_{r,j} - \overline{y}_r)/s_r$, for $j = 1, \ldots, 10$, should then be approximately standard normal variates. Because they share the terms $\overline{y}_r$ and s_r in

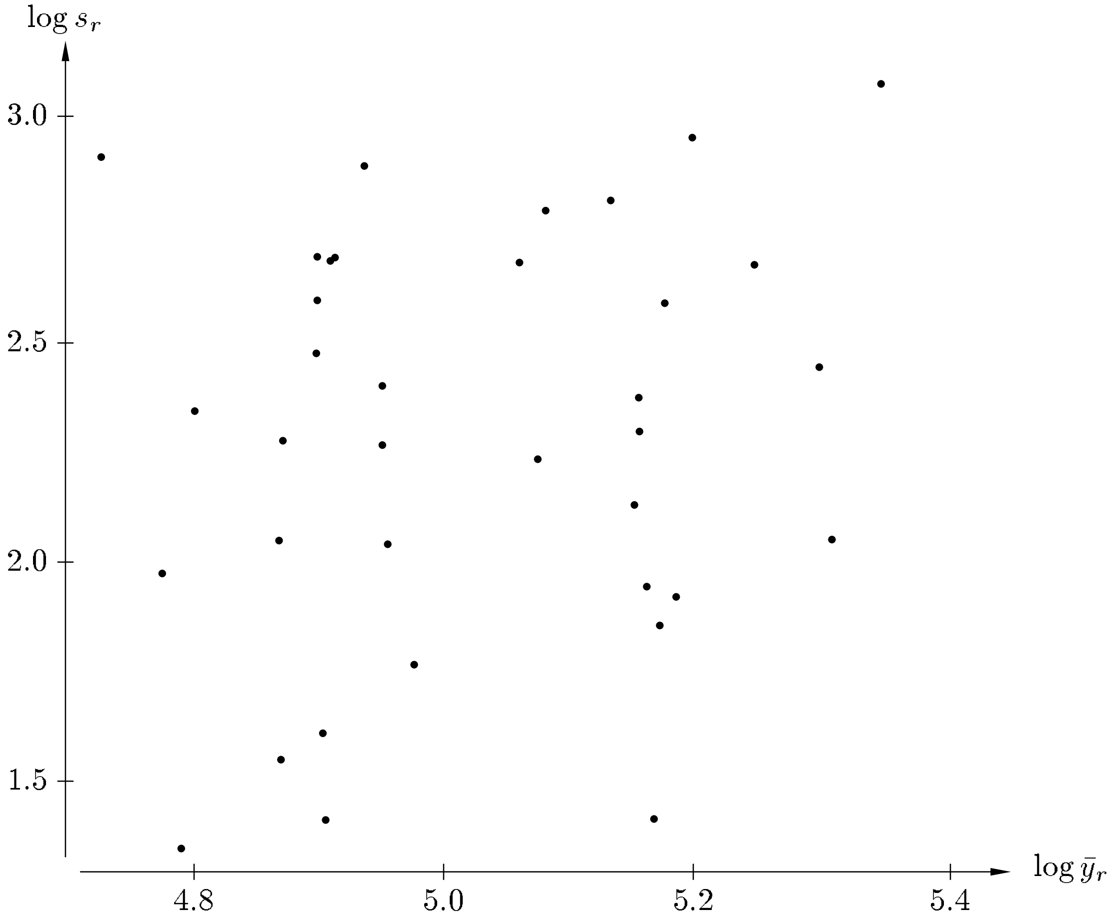

Fig. 3: Functional separation scatter plot. Proportionality of the standard deviation to a power of the mean, of the form $\sigma = \alpha\mu^\beta$ with $\beta \neq 0$, entails a nonzero correlation between the sample quantities $\log \bar{y}_r$ and $\log s_r$, for $r = 1, \ldots, 36$. The scatter plot provides no evidence for a nonzero correlation.

common, $x_{r,1}, \ldots, x_{r,10}$ are dependent random variables, but we did not follow up this lack of independence.

The normal plot in Figure 2 provides a graphical way of testing whether we see independent replications from a standard normal distribution. The plot shows the 360 pairs (x, y), where x is the kth largest of the 360 values $x_{r,j}$, and y is the $(k - 0.5)/360$ quantile of a standard normal distribution. If the numbers $x_{r,j}$ originate from a normal distribution, then they will be close to their corresponding quantiles, and the plot will show a fairly straight line. Indeed, in our case we obtained a correlation coefficient of 0.9973. Hence there is no indication that normality fails for the data set under discussion.

Another reason to contemplate a transformation of the data is functional separation, of the standard deviation σ from the mean μ. In many technical problems it is conceivable that the standard deviation grows proportionally to the mean or a power of the mean, of the form $\sigma = \alpha\mu^\beta$. Application of the logarithm turns this into a straight line relation, $\log \sigma = \log \alpha + \beta \log \mu$. For each run r, the observed quantities would thus follow the model $\log s_r = \log \alpha + \beta \log \overline{y}_r$. The corresponding scatter plot, in Figure 3, does not suggest a nonzero correlation between $\log s_r$ and $\log \overline{y}_r$. With $\beta = 0$, there is no need to transform the data. For more details the reader is referred to pages 285–286 in Box and Draper (1987), or pages 258–259 in Logothetis and Wynn (1989).

A third approach to improve the distributional assumptions uses a Box–Cox power transformation y^p, for some exponent $p \neq 0$ (and $\log y$ for $p = 0$). This class of transformations goes back to Box and Cox (1964), as described on pages 289–290 of Box and Draper (1987), or pages 255–256 of Logothetis and Wynn (1989). For our data the 95% confidence interval for p ran from 0.32 up to 1.14, thus including the value $p = 1$. This again indicated that the data did not call for a transformation.

8 Location Analysis

In order to estimate the parameters $\theta_0, \ldots, \theta_{22,22}$ in model (1) we took recourse to the analysis of variance. Coding the levels $-, 0, +$ into the numbers $-1, 0, +1$, we thus obtained the least squares estimates $\widehat{\theta}_i$ and $\widehat{\theta}_{i,i}$ for the parameters in model (1). The interpretation of the estimated values is complicated by the fact that the estimators are correlated. For interpreting the results, it is helpful to reparameterize model (1) in such a way that the estimators for the linear and quadratic terms become uncorrelated.

This type of parameter transformation corresponds to a decomposition of the sum of squares. To see this, consider any fixed three-level factor. Let $\widehat{\mu}_1, \widehat{\mu}_2, \widehat{\mu}_3$ be the average of all observations when the level is low, medium, or high. With the grand average $\widehat{\mu}$, we then have

$$\sum_{\ell=1}^{3} (\widehat{\mu}_\ell - \widehat{\mu})^2 = \frac{(\widehat{\mu}_3 - \widehat{\mu}_1)^2}{2} + \frac{(\widehat{\mu}_1 - 2\widehat{\mu}_2 + \widehat{\mu}_3)^2}{6}.$$

The first term on the right hand side gives the sum of squares associated with the linear effect, comparing the responses at the two extreme levels of the factor. The second term evaluates the quadratic effect, averaging the responses at the low and high level, relative to the one in-between.

The new parameters are defined by

$$\lambda_0 = \theta_0 + \frac{2}{3} \sum_{i=12}^{22} \theta_{i,i}, \qquad \lambda_i = \sqrt{3}\theta_i \qquad \text{for } i = 1, \ldots, 11,$$

$$\lambda_i = \sqrt{2}\theta_i, \qquad \lambda_{i,i} = \frac{\sqrt{6}}{3}\theta_{i,i} \qquad \text{for } i = 12, \ldots, 22.$$

With the new parameters, model (1) turns into the location model,

$$
\begin{aligned}
y = \lambda_0 &+ \frac{1}{\sqrt{3}} t_1 \lambda_1 + \cdots + \frac{1}{\sqrt{3}} t_{11} \lambda_{11} + \frac{1}{\sqrt{2}} t_{12} \lambda_{12} + \cdots + \frac{1}{\sqrt{2}} t_{22} \lambda_{22} \\
&+ \frac{1}{\sqrt{6}} (3t_{12}^2 - 2) \lambda_{12,12} + \cdots + \frac{1}{\sqrt{6}} (3t_{22}^2 - 2) \lambda_{22,22}.
\end{aligned} \tag{2}
$$

Based on the t-ratios, most of the estimated linear effects $\widehat{\lambda}_i$ and the estimated quadratic effects $\widehat{\lambda}_{i,i}$ turned out to be significantly distinct from zero. In other words, all parameters except a few ought to be included in model (2). However, we felt that the large number of significant effects was due to a considerable bias in the variance estimate, in that the true variance was severely underestimated.

As an alternative analysis we studied a normal plot of the parameter estimates $\widehat{\lambda}_i$ and $\widehat{\lambda}_{i,i}$. Assuming that the effects vanish, the estimates $\widehat{\lambda}_i$ and $\widehat{\lambda}_{i,i}$ have a standard normal distribution. Therefore the normal plot of the estimated effects should result in a straight line. For positive effects, the estimates will deviate towards large values and thus disclose themselves for identification, see Figure 4.

The normal plot drastically reduced the number of significant effects, and left us with $\widehat{\lambda}_5 = 20.79$, $\widehat{\lambda}_{15,15} = 16.41$, and $\widehat{\lambda}_{13} = 15.69$. Since these values are positive, maximization of the burst pressure in model (2) calls for the high level of factors t_5 and t_{13}. For the three-level factor t_{15}, either level high or low seems best in model (2). We chose the high level since the associated linear effect $\widehat{\lambda}_{15} = 6.94$ was positive.

The normal plot also affords a rough idea of the standard error of the least squares estimates, as the number by which the abscissa has to be scaled so that the fitted line has slope unity. Figure 4 yields a value of about 7, hence the graphical impression that the estimates $\widehat{\lambda}_5$, $\widehat{\lambda}_{13}$ and $\widehat{\lambda}_{15,15}$ are shifted in the direction of positive location effects is complemented by the numerical statement that the values 20.79, 16.41, and 15.69 are bigger than the upper 2σ point.

9 Dispersion Analysis

The goal of the dispersion analysis is to find out which factors have a significant influence on the variability of the manufacturing process. To this end we fitted a main-effects model to the logarithm of the sample standard deviations,

$$
\begin{aligned}
-\log s_r = \delta_0 &+ \frac{1}{\sqrt{3}} t_1 \delta_1 + \cdots + \frac{1}{\sqrt{3}} t_{11} \delta_{11} + \frac{1}{\sqrt{2}} t_{12} \delta_{12} + \cdots + \frac{1}{\sqrt{2}} t_{22} \delta_{22} \\
&+ \frac{1}{\sqrt{6}} (3t_{12}^2 - 2) \delta_{12,12} + \cdots + \frac{1}{\sqrt{6}} (3t_{22}^2 - 2) \delta_{22,22}.
\end{aligned} \tag{3}
$$

The multiplication of $\log s_r$ by -1 in (3) is negligible, but seemed to aid communication. In model (2), the aim was to *maximize* burst pressure. In the present model (3), we

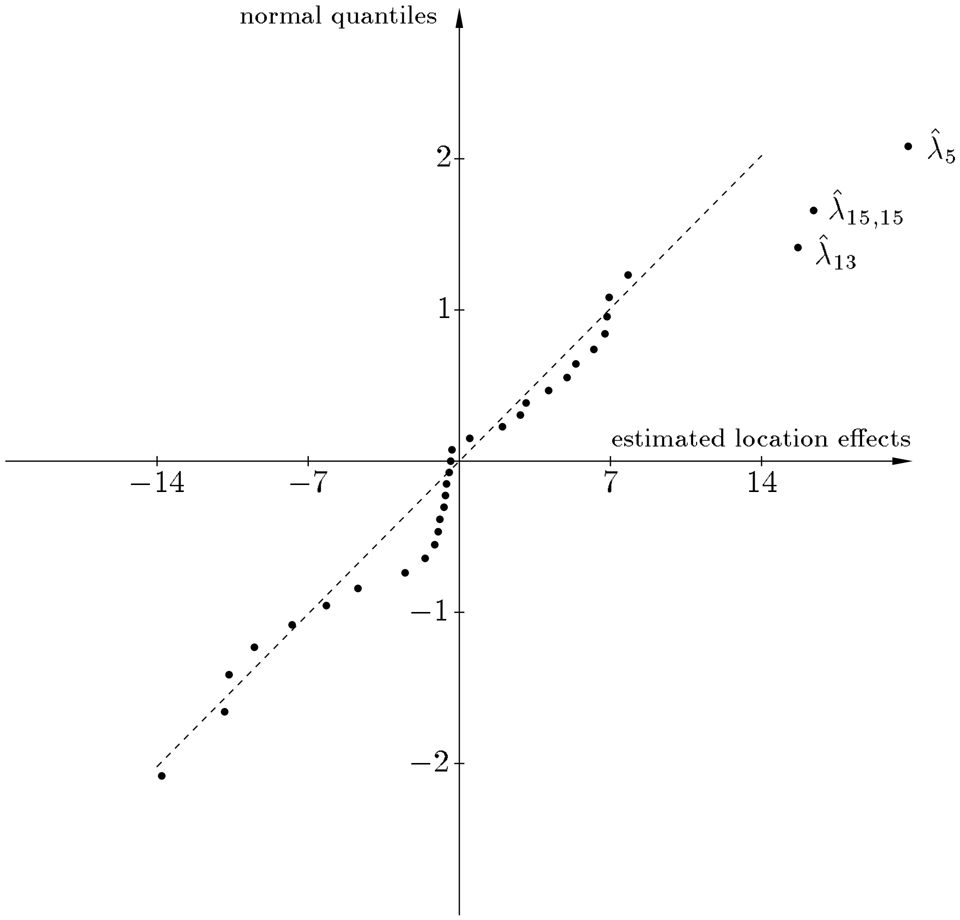

Fig. 4: Normal plot of estimated location effects. Under the null hypothesis of zero effects in the location model (2), the 33 scaled estimates $\hat\lambda_i$ and $\hat\lambda_{i,i}$ result in a straight line when ordered and plotted against the $(k-0.5)/33$ standard normal quantiles. The plot suggests that $\hat\lambda_5, \hat\lambda_{15,15}, \hat\lambda_{13}$ deviate from the null hypothesis.

study $-\log s_r = \log(1/s_r)$. Hence again the goal is one of *maximizing* the precision $1/s_r$, rather than to minimize the process standard deviation s_r.

For the dispersion model (3), 36 observations were available, one for each run. With 34 parameters for the mean, this leaves only 2 degrees of freedom for estimating the error variance. We found this insufficient to rely on the classical analysis of variance table. Again we resorted to a normal plot, see Figure 5.

The significant dispersion effects appeared to be $\widehat{\delta}_6 = 0.38$, $\widehat{\delta}_{10} = 0.30$, and perhaps also $\widehat{\delta}_{18} = 0.225$. Since these are positive, the high levels for factors t_6, t_{10}, and t_{18} are the appropriate choices to lead to a small standard deviation.

Furthermore Figure 5 suggests that the least squares estimates of the parameters in model (3) have a standard error in the vicinity of 0.125. With this estimate, $\widehat{\delta}_6$ and $\widehat{\delta}_{10}$ are clearly beyond the 2σ point. However, we decided that no harm is done to also include $\widehat{\delta}_{18}$ as a dispersion effect.

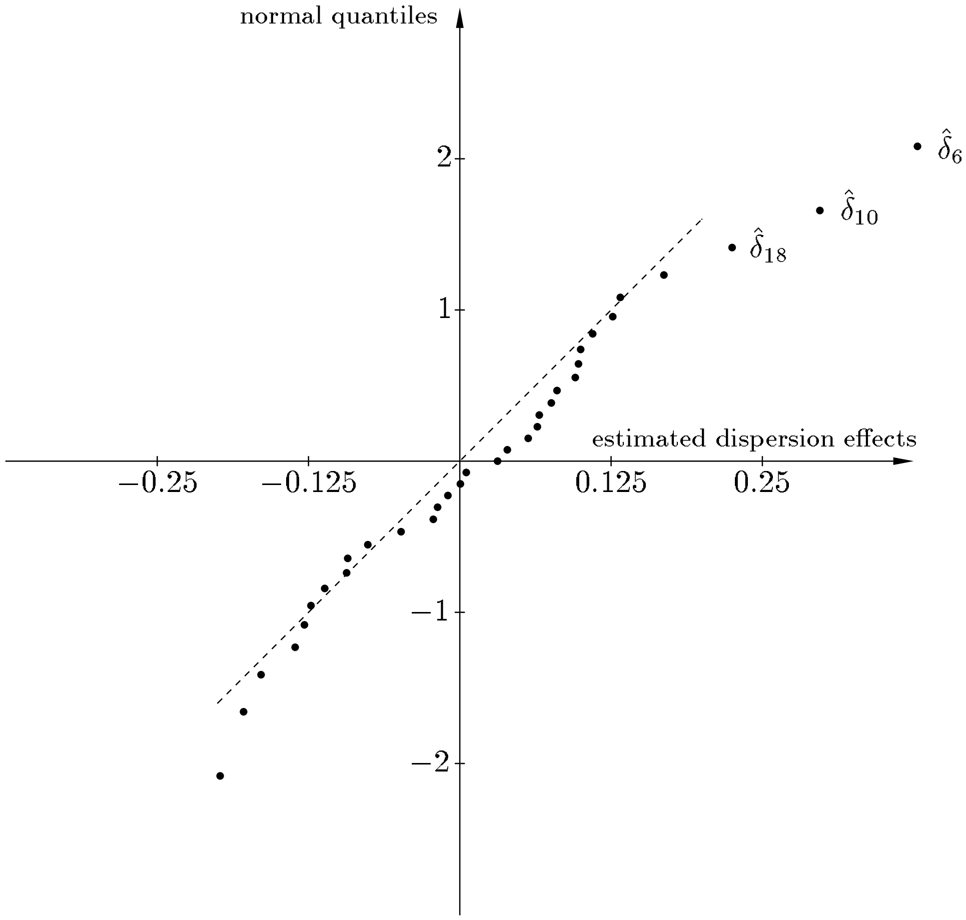

Fig. 5: Normal plot of estimated dispersion effects. Assuming a dispersion model (3) with vanishing effects, the 33 estimates $\hat{\delta}_i$ and $\hat{\delta}_{i,i}$ will give a straight line when ordered and plotted against the $(k-0.5)/33$ standard normal quantiles. The shift of $\hat{\delta}_6, \hat{\delta}_{10}, \hat{\delta}_{18}$ to the right from the straight line indicates positive effects.

10 Factor Classification

Merging the results from the location analysis and from the dispersion analysis, we concluded that t_5, t_{13}, and t_{15} are signal factors, while t_6, t_{10}, and t_{18} are control factors. All other factors were treated as nuisance factors.

A synthesis of the location analysis and the dispersion analysis is graphically shown in the location–dispersion diagram of Figure 6. The (empty) top right rectangle features the effects that are significant in either analyses. The bottom right, and the top left rectangles contain the effects that are significant in the location analysis and in the dispersion analysis, respectively. The nonsignificant effects appear in the bottom left rectangle.

The location dispersion diagram is a convenient tool to communicate to the engineers the factor classification that flows from the two analyses. In Figure 6, the dashed lines have been drawn through the bounding nonsignificant effects, so that the

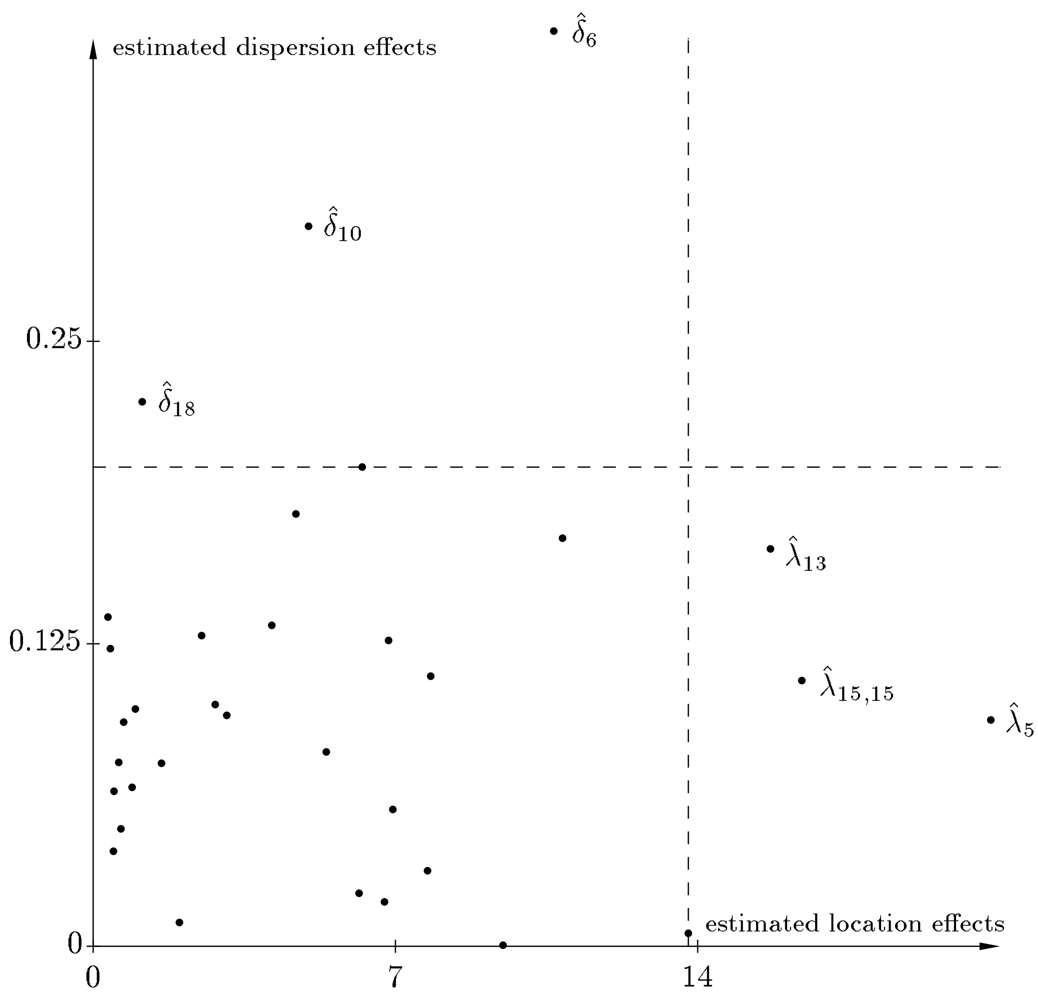

Fig. 6: Location–dispersion diagram. The diagram shows the 33 pairs of absolute values of estimated effects in the location model (2) and the dispersion model (3), $(|\lambda_i|, |\delta_i|)$ and $(|\lambda_{i,i}|, |\delta_{i,i}|)$. Significant estimates for location effects stand out to the right; significant dispersion effects are at the top.

significant effects stand out as clearly as possible. The diagram is a reduced adaption of the location–dispersion plot presented in Nair and Pregibon (1986).

We concluded that setting the factors t_5, t_6, t_{10}, t_{13}, t_{15}, t_{18} on the high level would maximize burst pressure. At the same time these settings would maximize the precision of the process, that is, minimize process variability.

For the purpose of prediction, we set all other factors to the level for which burst pressure is maximized. While the other factors are nonsignificant in the normal plot analysis, most of them appear to be significant in the analysis of variance evaluation of model (2). In any case, we had to tell the operators what levels to choose, and we decided that the levels might just as well be chosen to optimize burst pressure. The confidence interval for the mean burst pressure under this setting ranged from 198 to 212 bar.

11 Confirmation Experiment

The series of investigation terminated with an implementation of the recommended settings for the factor levels. The observed data 181, 230, 185, 229, 196, 210, 215, 193, 202, and 223 averaged to 206.4 bar, and thus conformed well with the model prediction. However, the observed standard deviation of 17.8 bar was much higher than the model suggested. In any case, relative to the standard operating conditions of run *25*, the improvement was spectacular.

Of course, from the practical point of view it is totally irrelevant whether the statistical model is "correct", and whether any theoretical assumptions are satisfied or not. The important success for the engineers who were responsible for setting up the production process, and for the operators who were responsible for running it, was the improvement that was actually achieved. The confirmation experiment well documented that improvement, and also set a formal end to the cooperation.

12 Discussion

There are alternative analyses that one may contemplate, and we carried out some of them. For instance, we could not find any significant effect associated with the sequencing in which the runs were realized. Also the experiment was carried out in sections, over a time span of 9 weeks. We did not find that weeks led to any significant block effect.

The major problem with the present data was that they called for a weighted least squares analysis, since the per run variability in Figure 1 is evidently seen to be nonconstant. However, this more advanced analysis did not lead to qualitatively different conclusions. Therefore we decided to prefer the simpler method, at the expense of theoretical justifications, but with the aadvantage to make the communication with the engineers easier.

Evidently the experiment was a major enterprise within the company. Generally it is a bad strategy to run just one very large experiment, as it invariably tends to leave many questions open and to be exhaustive only in terms of the stamina of all concerned. It is much better to run a number of small experiments, which then lead to a continuous, consecutive improvement of the process.

The optimal factor settings were distinct from the settings in run *7* that was best among the *36* runs. Hence a "pick-the-winner" rule appears to be a rather naïve strategy. Moreover, with smaller experiments and fewer runs, the pick-the-winner rule may miss the optimal settings even more severely.

Nevertheless it is remarkable that the total number of possible combinations of the 11 two-level and 11 three-level factors exceeds 360 million. Only a tiny fraction, 36 combinations, were observed in the experiment. Clearly, these 36 runs must be designed with care in order to give rise to a data set that carries information of any conclusive value.

References

Abt, M. and Pukelsheim, F. (1995). "Improving manufacturing quality through planned experiments: Statistical methodology." *Surveys on Mathematics for Industry* **5** 27–33.

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. (1987). *Empirical Model-Building and Response Surfaces.* Wiley, New York.

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

Nair, V.N. and Pregibon, D. (1986). "A data analysis strategy for quality engineering experiments." *AT&T Technical Journal* **65** 73–84.

Taguchi, G. (1960). "Tables of orthogonal arrays and linear graphs." *Report of Statistical Application Research, Union of Japanese Scientists and Engineers* **7** 1–52.

Weihs, C., Berres, M. and Grize, Y.-L. (1995). "Statistical design of experiments in industrial practice." *Surveys on Mathematics for Industry* **5** 49–73.

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

Communicated by M. Deistler