PETER GOOS
BRADLEY JONES

OPTIMAL DESIGN OF EXPERIMENTS

A Case Study Approach

WILEY

7

A response surface design in blocks

7.1 Key concepts

1. When it is possible to group the runs of an experiment in such a way that runs in each group are more like each other than they are like runs in a different group, then it is statistically more efficient to make this grouping explicit. This procedure is called blocking the experiment.
2. It is important to identify the characteristic that makes the runs in a group alike. For example, runs done in one day of processing are generally more alike than those performed on separate days. In this case, the blocking factor is day, and runs performed on a given day constitute one block.
3. An important characteristic of a blocking factor is that such factors are not under the direct control of the investigator. For example, you can choose to include different lots of material from a supplier in your experiment, but you cannot reproduce any lot in the future.
4. It is ideal to make the blocking factor orthogonal to the other factors. Real-world constraints on model form, block sizes and the total number of runs can make this impossible. In such cases, optimal blocking is a useful alternative.
5. For making inferences about future blocks, it can be useful to model the characteristics that make one block different from another as random variation.
6. Characterizing the impact of the blocking factor as random variation requires the use of a mixed model (having both fixed and random components). Mixed models are better fit using generalized least squares (GLS) rather than ordinary least squares (OLS).

Optimal Design of Experiments: A Case Study Approach, First Edition. Peter Goos and Bradley Jones.

When necessity dictates that you perform an experiment in groups of runs, we say that you have blocked the experiment. The blocking factor is the condition that defines the grouping of the runs. In our case study, the blocking factor is day, because the process mean shifts randomly from day to day and the investigator wants to explicitly account for these shifts in the statistical model.

There is more than one way to construct a blocked experiment. We discuss this, make comparisons, and recommend a general approach for the design and analysis of blocked experiments.

7.2 Case: the pastry dough experiment

7.2.1 Problem and design

At the reception on the evening of the opening day of the annual conference of the European Network for Business and Industrial Statistics, Peter and Brad are discussing the day's presentations on a rooftop terrace with a view on the Acropolis in Athens, Greece. One of the Greek conference attendees walks up to them.

[Maroussa] Hi. I am Maroussa Markianidou. I am a researcher in the Department of Food Science and Technology at the University of Reading in England.

[Peter] Nice to meet you, Maroussa. I am Peter, and this is Brad.

The two men and the lady shake hands and exchange business cards, after which Maroussa Markianidou continues the conversation.

[Maroussa, addressing Peter] I saw your presentation on the optimal design of experiments in the presence of covariates this afternoon. I liked the flexibility of the approach, and wondered whether it would be suitable for an experiment I am involved in.

[Peter] What kind of experiment are you doing?

[Maroussa] We're still in the planning phase. We decided to run an experiment at our lab in Reading about the baking of pastry dough to figure out the optimal settings for a pilot plant production process. It looks as if we are going to study three factors in the experiment, and we will measure two continuous responses that are related to the way in which the pastry rises.

[Peter] What kind of factors will you use?

[Maroussa] There are three continuous factors: the initial moisture content of the dough, the screw speed of the mixer, and the feed flow rate of water being added to the mix. We plan to study these factors at three levels because we expect substantial curvature. Therefore, we are considering a modified central composite design.

[Brad] Why a *modified* central composite design? What is wrong with the standard central composite design?

[Maroussa] Well, the experiment will require several days of work, and our experience with baking pastry is that there is quite a bit of day-to-day variation. Of course, we do not want the day-to-day variation to have an impact on our conclusions. So, we need to find a way to run the central composite design so that the effects of our three factors are not confounded with the day effects.

Table 7.1 Orthogonally blocked three-factor central composite design with two blocks of six runs and one block of eight runs.

Block	x_1	x_2	x_3
1	−1	−1	−1
	−1	1	1
	1	−1	1
	1	1	−1
	0	0	0
	0	0	0
2	−1	−1	1
	−1	1	−1
	1	−1	−1
	1	1	1
	0	0	0
	0	0	0
3	−1.633	0	0
	1.633	0	0
	0	−1.633	0
	0	1.633	0
	0	0	−1.633
	0	0	1.633
	0	0	0
	0	0	0

[Brad] I see. Please continue.

[Maroussa] I know that there are textbooks on experimental design and response surface methodology that deal with the issue of running central composite designs in groups, or blocks, so that the factor effects are not confounded with the block effects. The problem with the solutions in the textbooks is that they don't fit our problem. Look at this design.

Maroussa digs a few stapled sheets of paper out of her conference bag, and shows the design given in Table 7.1. In the table, the three factors are labeled x_1, x_2, and x_3, and the factor levels appear in coded form.

[Maroussa] According to the textbooks, this is *the* way to run a three-factor central composite design in blocks. The design has three blocks. The first two blocks contain a half fraction of the factorial portion of the central composite design plus two center runs. The third block contains the axial points plus two center runs.

[Peter] Why is it that this design doesn't fit your problem?

[Maroussa] There are several reasons. First, we can only do four runs a day. So, for our problem, we need blocks involving four runs only. Second, we have scheduled the laboratory for seven days in the next couple of weeks, so that we need a design with seven blocks. The third reason why I don't like the design is the fact that the third block involves the factor levels ± 1.633. Apparently, the value 1.633 is required to ensure that the design is what the textbooks call orthogonally blocked. You know better than me that this is jargon for saying that the factor effects are not confounded with the block effects. However, we would much rather use just three levels for every factor. The problem is that using ± 1 for the nonzero factor levels in the axial points violates the condition for having an orthogonally blocked design.

[Peter] Using ± 1 instead of ± 1.633 for the axial points would not destroy all of the orthogonality. The main effects and the two-factor interaction effects would still be unconfounded with the block effects, but the quadratic effects would not. So, you could then estimate the main effects and the two-factor interaction effects independently of the block effects, but not the quadratic effects.

[Maroussa, nodding] I see. Anyway, I was thinking that we needed a different approach. Since we can explore four factor-level combinations on each of seven days, we need a design with 28 runs. The standard central composite design for a three-factor experiment involves eight factorial points, six axial points, and a couple of center runs. I was thinking that a good start for our design problem would be to duplicate the factorial points and to use six replicates of the center point. This would give us 28 runs: the duplicated factorial portion would give us 16 runs, and we would have six center runs plus six axial runs.

[Brad] And how are you going to arrange these 28 runs in blocks of four?

[Maroussa] This is where I got stuck. I was unable to figure out a way to do it. It is easy to arrange the factorial points in four blocks of size four. That kind of thing is described in the textbooks too.

Maroussa flips a few pages, and shows the design in Table 7.2. The two replicates of the factorial portion of the central composite design are blocked by assigning the points with a $+1$ for the third-order interaction contrast column $x_1x_2x_3$ to one

Table 7.2 Duplicated two-level factorial design arranged in four orthogonal blocks of four runs using the contrast column $x_1x_2x_3$.

Replicate 1					Replicate 2				
Block	x_1	x_2	x_3	$x_1x_2x_3$	Block	x_1	x_2	x_3	$x_1x_2x_3$
1	1	1	−1	−1	3	1	1	−1	−1
	−1	−1	−1	−1		−1	−1	−1	−1
	−1	1	1	−1		−1	1	1	−1
	1	−1	1	−1		1	−1	1	−1
2	−1	−1	1	1	4	−1	−1	1	1
	1	−1	−1	1		1	−1	−1	1
	−1	1	−1	1		−1	1	−1	1
	1	1	1	1		1	1	1	1

block and those with a −1 for that column to another block. This ensures that the main effects and the two-factor interaction effects are not confounded with the block effects.

[Maroussa] The problem is that I don't find guidance anywhere as to how to arrange the axial points and the center runs in blocks. The only thing I found is that one sentence in my textbook on response surface methodology that says that it is crucial that all the axial points are in one block.

[Peter, grinning] That advice is pretty hard to follow if you have six axial points and you can only manage four runs in a block.

[Maroussa] Exactly. Do you have suggestions to help us out?

[Brad] My suggestion would be to forget about the central composite design. For the reasons you just mentioned, it is not a fit for your problem. My advice would be to match your design directly to your problem. So, you want 28 factor-level combinations arranged in seven blocks of size four that allow you to estimate the factor effects properly.

[Maroussa] What do you mean by *properly*?

[Brad] Ideally, independently of the day effects. If that is impossible, as precisely as possible.

[Maroussa] And how do you do that?

[Brad, taking his laptop out of his backpack] It is easy enough with the optimal design algorithms that are available in commercial software packages these days. The packages allow you to specify the blocking structure and then produce a statistically efficient design with that structure in the blink of an eye.

While Brad's laptop is starting up, Peter chimes in.

[Peter] Yeah, all you have to do is specify the number of factors, their nature, the model you would like to fit, the number of blocks, and the number of runs that can be done within a block. Brad will show you.

Brad, who is now ready for the demonstration, takes over.

[Brad] Can you remind me what the three factors in your experiment were? They were continuous, weren't they?

[Maroussa, flipping through the pages in front of her and finding one showing Table 7.3] Sure, here they are.

[Brad] Wow. You came prepared.

Brad inputs the data in the user interface of the software he is using.

[Peter] Is there a problem with randomizing the order of the runs within each day and resetting the factor levels for every run? Or would that be inconvenient or costly?

Table 7.3 Factors and factor levels used in the pastry dough mixing experiment.

Flow rate (kg/h)	Moisture content (%)	Screw speed (rpm)
30.0	18	300
37.5	21	350
45.0	24	400

[Maroussa] As far as I can see, any order of the runs is equally convenient or inconvenient.

[Brad, pointing to the screen of his laptop] This is where I specify that we want to fit a response surface model in the three factors. That means we would like to estimate the main effects of the factors, their two-factor interaction effects, and their quadratic effects. Next, I specify that we need 28 runs in seven blocks of four runs, and then all I have to do is put the computer to work.

Brad clicks on a button in his software, and, a few second later, the design in Table 7.4 pops up.

[Brad] I have a design for you.

[Maroussa] Gosh, I haven't even had time to sip my drink. How can I see now that this is the design we need back in Reading?

[Brad] The key feature of the design is that it matches your problem. We have seven blocks of four runs. Also, you can see that the design has three levels for each of the factors. For instance, the moisture content in the first run on the first day is 30%, which is the low level of that factor. In the second run, the moisture content is 45%, which is the high level. In the third run, the moisture content is right in the middle, at 37.5%. That we have three levels for every factor can be better seen by switching to coded factor levels.

Brad enters the formulas to transform the factor levels in the design in Table 7.4 into coded form:

$$x_1 = (\text{flow rate} - 37.5)/7.5,$$
$$x_2 = (\text{moisture content} - 21)/3,$$

and

$$x_3 = (\text{screw speed} - 350)/50.$$

Next to the design in uncoded form, the coded design in Table 7.5 appears, and Brad continues.

[Brad] Here you can easily see that we have three levels for every factor: a low, a high, and an intermediate level. This is what you wanted, isn't it?

[Maroussa, pointing to Table 7.1] Right. Now, how can I see that the design you generated is orthogonally blocked, just like the central composite design here.

[Peter] It isn't.

[Maroussa, looking skeptical] I beg your pardon?

[Peter] In many practical situations, it is impossible to create an orthogonally blocked response surface design. This is certainly so if the number of runs within each block is small, and when you have to fit a response surface model. Constructing an orthogonally blocked design is easier to do for two-level factorial and fractional factorial designs, when the number of runs within the blocks is a power of two and your model is not too complicated.

Table 7.4 D-optimal design for the pastry dough experiment in engineering units.

Day	Flow rate (kg/h)	Moisture content (%)	Screw speed (rpm)
1	30.0	18	300
	45.0	18	400
	37.5	21	350
	45.0	24	300
2	37.5	24	300
	45.0	24	400
	45.0	18	350
	30.0	21	400
3	37.5	18	300
	45.0	21	400
	45.0	24	300
	30.0	24	350
4	30.0	18	300
	45.0	18	400
	30.0	24	400
	37.5	21	350
5	45.0	24	350
	30.0	21	300
	37.5	18	400
	30.0	24	400
6	30.0	18	400
	45.0	18	300
	45.0	24	400
	30.0	24	300
7	30.0	18	350
	37.5	24	400
	30.0	24	300
	45.0	21	300

[Maroussa, nodding] But this means that the day-to-day variation in our process will influence the estimates of the factor effects. Doesn't that invalidate any conclusion I draw from the data?

[Peter] No, it's not that bad. The design Brad generated is a D-optimal block design. There is a link between the D-optimality of block designs and orthogonality. You

Table 7.5 D-optimal design for the pastry dough experiment in coded units.

Day	Flow rate x_1	Moisture content x_2	Screw speed x_3
1	−1	−1	−1
	1	−1	1
	0	0	0
	1	1	−1
2	0	1	−1
	1	1	1
	1	−1	0
	−1	0	1
3	0	−1	−1
	1	0	1
	1	1	−1
	−1	1	0
4	−1	−1	−1
	1	−1	1
	−1	1	1
	0	0	0
5	1	1	0
	−1	0	−1
	0	−1	1
	−1	1	1
6	−1	−1	1
	1	−1	−1
	1	1	1
	−1	1	−1
7	−1	−1	0
	0	1	1
	−1	1	−1
	1	0	−1

could say that, in a way, D-optimal block designs are as close to being orthogonally blocked as the number of blocks and the number of runs in each block allow.

[Maroussa, all ears] As close as possible to being orthogonally blocked might still not be close enough.

[Peter] My experience is that D-optimal block designs are close enough to being orthogonally blocked that you do not need to worry about the confounding between the factor effects and the block effects. There are several things you can do to quantify to what extent a block design is orthogonally blocked. One approach is to compute an efficiency factor for each factor effect in your model. The minimum and maximum values for the efficiency factors are 0% and 100%. An efficiency factor of 100% is, of course, desirable. Designs that are perfectly orthogonally blocked have efficiency factors of 100%. An efficiency factor of 0% results when a factor effect is completely confounded with the block effects. In that case, you can't estimate the factor effect. However, as soon as an efficiency factor is larger than 0%, you can estimate the corresponding factor effect. The larger the efficiency factor, the more precisely you can estimate it.

[Maroussa, pointing to Brad's laptop] And what would the efficiency factors for this design be?

[Peter] I trust that they are larger than 90%, which would indicate that the design is very nearly orthogonally blocked. So, I am pretty sure there is no reason to worry about confounding between factor effects and block effects when using this design. If you would like to be fully sure, I can compute the values of the efficiency factors later this evening and let you know what they are tomorrow.

[Maroussa, looking persuaded and taking out a memory stick] That would be great. Could you put that design on my memory stick?

[Brad, copying the design into a spreadsheet and saving it onto Maroussa's memory stick] Sure. Here you go.

That evening, in his hotel room, Peter confirms that the efficiency factors for the D-optimal design with seven blocks of four runs range from 93.4% to 98.3%, with an average of 96.1%. The efficiency factors for the three main effects, the three interaction effects, and the three quadratic effects are shown in Table 7.6, along with the corresponding variance inflation factors (VIFs). The VIFs show to what extent the variances of the parameter estimates get bigger due to the blocking of the experiment. The efficiency factors and the VIFs are the reciprocals of each other. Both are based on the variances of the factor-effect estimates in the presence and in the absence of blocking.

Table 7.6 Efficiency factors and variance inflation factors (VIFs) for the D-optimal design for the pastry dough mixing experiment, as well as the variances of the factor-effect estimates in the presence and in the absence of blocking.

Effect	Blocking	No blocking	Efficiency (%)	VIF
Flow rate	0.0505	0.0471	93.4	1.0705
Moisture content	0.0495	0.0465	94.1	1.0632
Screw speed	0.0505	0.0471	93.4	1.0705
Flow rate × Moisture content	0.0595	0.0579	97.3	1.0274
Flow rate × Screw speed	0.0580	0.0570	98.3	1.0178
Moisture content × Screw speed	0.0595	0.0579	97.3	1.0274
Flow rate × Flow rate	0.2282	0.2215	97.0	1.0305
Moisture content × Moisture content	0.2282	0.2215	97.0	1.0305
Screw speed × Screw speed	0.2282	0.2215	97.0	1.0305

7.2.2 Data analysis

A couple of weeks later, Brad enters Intrepid Stats's European office, where Peter is working on his desktop computer.

[Peter, handing Brad a sheet of paper containing Table 7.7]. You remember Maroussa Markianidou, the Greek lady from Reading in the UK? She ran the design

Table 7.7 Design and responses for the pastry dough experiment. The responses are the longitudinal (y_1) and cross-sectional (y_2) expansion indices.

Day	Flow rate (kg/h)	Moisture content (%)	Screw speed (rpm)	x_1	x_2	x_3	y_1	y_2
1	30.0	18	300	−1	−1	−1	15.6	5.5
	45.0	18	400	1	−1	1	15.7	6.4
	37.5	21	350	0	0	0	11.2	4.8
	45.0	24	300	1	1	−1	13.4	3.5
2	37.5	24	300	0	1	−1	13.6	4.7
	45.0	24	400	1	1	1	15.9	5.5
	45.0	18	350	1	−1	0	16.2	6.3
	30.0	21	400	−1	0	1	13.2	5.0
3	37.5	18	300	0	−1	−1	12.5	4.5
	45.0	21	400	1	0	1	14.0	4.1
	45.0	24	300	1	1	−1	11.9	4.3
	30.0	24	350	−1	1	0	9.2	4.3
4	30.0	18	300	−1	−1	−1	13.7	5.5
	45.0	18	400	1	−1	1	17.2	6.0
	30.0	24	400	−1	1	1	11.7	4.5
	37.5	21	350	0	0	0	11.9	4.9
5	45.0	24	350	1	1	0	10.9	4.1
	30.0	21	300	−1	0	−1	9.5	3.5
	37.5	18	400	0	−1	1	14.7	5.7
	30.0	24	400	−1	1	1	11.9	4.6
6	30.0	18	400	−1	−1	1	15.2	6.1
	45.0	18	300	1	−1	−1	16.0	5.1
	45.0	24	400	1	1	1	15.1	4.8
	30.0	24	300	−1	1	−1	9.1	4.4
7	30.0	18	350	−1	−1	0	15.6	6.6
	37.5	24	400	0	1	1	14.0	4.9
	30.0	24	300	−1	1	−1	11.6	4.8
	45.0	21	300	1	0	−1	14.2	4.5

that you generated for her in Athens. She e-mailed us this morning to ask whether we can analyze the data for her.

[Brad] Great. Please tell me she is going to pay us.

[Peter] Nope. I already replied to her that I was prepared to analyze her data if she agreed to put our names on her paper and if she shares the details about the application with us so that we can use her design problem as a case study in our book on modern experimental design. She skyped me soon after that to conclude the deal and to give me some of the background of the experiment.

[Brad] Fine. We need a chapter on blocking response surface designs.

Several days later, Peter sends Maroussa an e-mail report on the analysis of the data from her pastry dough experiment. Here is his e-mail:

> Dear Maroussa: I analyzed the data from the pastry dough experiment you conducted to study the impact of the moisture content of the dough, the screw speed of the mixer and the feed flow rate on the two responses you measured, the longitudinal expansion index (y_1, expressed in centimeters per gram) and the cross-sectional expansion index (y_2, also expressed in centimeters per gram). I also determined factor settings that will give you the desired target values of 12 cm/g and 4 cm/g for the longitudinal and the cross-sectional expansion index, respectively.
>
> The design that you ran was a D-optimal design with seven blocks of four runs, because you could do only four runs on each of seven days. Since you expected curvature, we generated the design assuming that the following second-order response surface model was adequate:

$$Y_{ij} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{kij} + \sum_{k=1}^{2} \sum_{l=k+1}^{3} \beta_{kl} x_{kij} x_{lij} + \sum_{i=1}^{3} \beta_{kk} x_{kij}^2 + \gamma_i + \varepsilon_{ij}, \tag{7.1}$$

> where x_{kij} is the coded level of the kth factor at the jth run on day i (i goes from 1 to 7, and j goes from 1 to 4), β_0 denotes the intercept, β_k is the main effect of the kth factor, β_{kl} is the effect of the interaction involving the kth and the lth factor, and β_{kk} is the quadratic effect of the kth factor. The term γ_i represents the ith day (or block) effect, and ε_{ij} is the residual error associated with the jth run on day i. The γ_i in the model captures the day-to-day variation in the responses.
>
> Fitting the second-order response surface model in Equation (7.1) to the two responses in Table 7.7 using generalized least squares (GLS) and restricted maximum likelihood (REML) estimation gave me the parameter estimates, standard errors and p values shown in Table 7.8. The left panel of the table shows the results for the longitudinal expansion index, while the right panel contains the results for the cross-sectional expansion index. For each of these responses, several p values are substantially larger than 5%, indicating that you can simplify the models. Using stepwise backward elimination, I got the models in Table 7.9.

Table 7.8 Factor-effect estimates, standard errors and p values obtained for the two responses, longitudinal expansion index and cross-sectional expansion index, in the pastry dough experiment when fitting the full second-order response surface model in Equation (7.1).

	Longitudinal expansion index			Cross-sectional expansion index		
Effect	Estimate	Standard error	p Value	Estimate	Standard error	p Value
β_0	11.38	0.66	< .0001	4.66	0.25	< .0001
β_1	1.01	0.21	0.0003	−0.05	0.08	0.5159
β_2	−1.47	0.20	< .0001	−0.61	0.08	< .0001
β_3	0.73	0.21	0.0040	0.35	0.08	0.0008
β_{12}	0.42	0.22	0.0822	0.03	0.09	0.7329
β_{13}	−0.07	0.22	0.7565	0.09	0.09	0.3391
β_{23}	0.21	0.22	0.3639	−0.11	0.09	0.2474
β_{11}	0.33	0.44	0.4692	0.04	0.17	0.8028
β_{22}	1.30	0.44	0.0114	0.72	0.17	0.0011
β_{33}	1.05	0.44	0.0331	−0.33	0.17	0.0738

The indices 1, 2 and 3 in the first column of the table refer to the flow rate, the moisture content and the screw speed, respectively.

For the longitudinal expansion index, the main effects of the three experimental factors, moisture content, screw speed and flow rate, have statistically significant effects. Also, the interaction effect involving moisture content and flow rate is significant, as well as the quadratic effects of the moisture content and the screw speed. For the cross-sectional

Table 7.9 Factor-effect estimates, standard errors, and p values obtained for the two responses, longitudinal expansion index and cross-sectional expansion index, in the pastry dough experiment when fitting simplified models.

	Longitudinal expansion index			Cross-sectional expansion index		
Effect	Estimate	Standard error	p Value	Estimate	Standard error	p Value
β_0	11.59	0.58	< .0001	4.46	0.19	< .0001
β_1	0.99	0.20	0.0001			
β_2	−1.46	0.19	< .0001	−0.62	0.08	< .0001
β_3	0.75	0.20	0.0015	0.33	0.08	0.0005
β_{12}	0.46	0.21	0.0463			
β_{22}	1.33	0.41	0.0058	0.69	0.17	0.0007
β_{33}	1.08	0.41	0.0197			

The indices 1, 2 and 3 in the first column of the table refer to the flow rate, the moisture content and the screw speed, respectively.

expansion index, only the main effects of the factors moisture content and screw speed are significant, as well as the quadratic effect of the moisture content.

I then used the simplified models in Table 7.9,

$$y_1 = 11.59 + 0.99x_1 - 1.46x_2 + 0.75x_3 + 0.46x_1x_2 + 1.33x_2^2 + 1.08x_3^2,$$

for the longitudinal expansion index, and

$$y_2 = 4.46 - 0.62x_2 + 0.33x_3 + 0.69x_2^2,$$

for the cross-sectional expansion index, to find factor settings that give you the desired target values, 12 cm/g and 4 cm/g, for these expansion indices. If you prefer to think in engineering units, then you might prefer this way of writing the two models:

$$\begin{aligned}\begin{array}{c}\text{Longitudinal}\\ \text{expansion index}\end{array} &= 11.59 + 0.99\,\frac{\text{Flow rate} - 37.5}{7.5}\\ &\quad - 1.46\,\frac{\text{Moisture content} - 21}{3} + 0.75\,\frac{\text{Screw speed} - 350}{50}\\ &\quad + 0.46\,\frac{\text{Flow rate} - 37.5}{7.5}\;\frac{\text{Moisture content} - 21}{3}\\ &\quad + 1.33\left(\frac{\text{Moisture content} - 21}{3}\right)^2\\ &\quad + 1.08\left(\frac{\text{Screw speed} - 350}{50}\right)^2,\end{aligned}$$

and

$$\begin{aligned}\begin{array}{c}\text{Cross-sectional}\\ \text{expansion index}\end{array} &= 4.46 - 0.62\,\frac{\text{Moisture content} - 21}{3}\\ &\quad + 0.33\,\frac{\text{Screw speed} - 350}{50}\\ &\quad + 0.69\left(\frac{\text{Moisture content} - 21}{3}\right)^2.\end{aligned}$$

Setting the flow rate to 40.6 kg/h, the moisture content to 22.3% and the screw speed of the mixer to 300 rpm gets you right on target. You can see this from Figure 7.1, which plots the dependence of your two responses on the three experimental factors. In the figure, the vertical dashed lines give the optimal factor levels for your process. The horizontal dashed lines indicate the corresponding predicted values for the expansion indices and for the overall desirability.

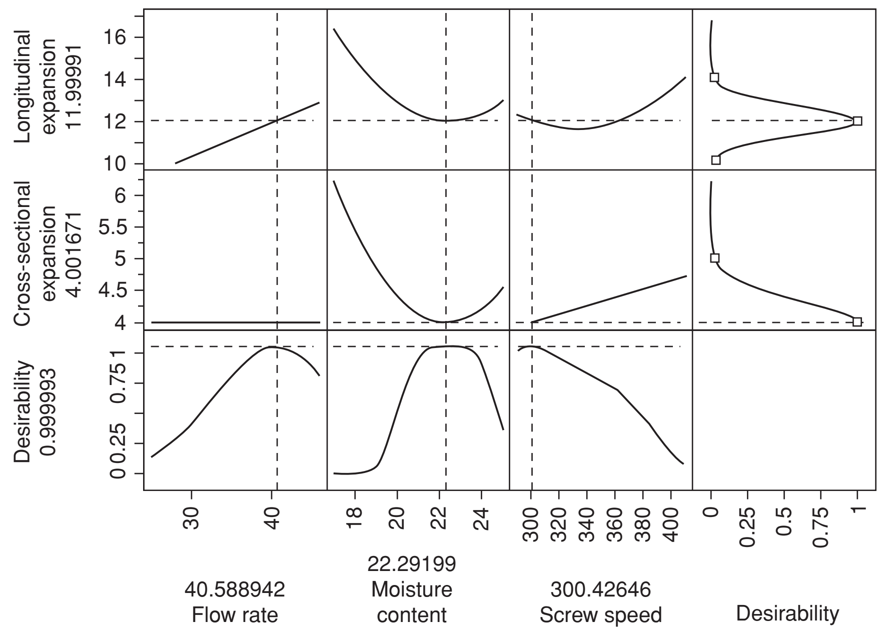

Figure 7.1 Plot of the factor-effect estimates, the desirability functions, and the optimal factor settings for the pastry dough experiment.

The upper panels of the figure show that the longitudinal expansion index linearly depends on the flow rate, and that it depends on the moisture content and the screw speed in a quadratic fashion. The middle panels show that the flow rate has no impact on the cross-sectional expansion index, that the screw speed has a linear effect, and that there is a quadratic relationship between the moisture content and the cross-sectional expansion index. The lower panels of the figure show how the overall desirability of the solution depends on the levels of the experimental factors. The overall desirability is a summary measure that tells us how close to target we are with our two responses. More specifically, it is the geometric mean of the individual desirabilities for your expansion indices, which are shown in the panels at the far right of Figure 7.1. You can see that the individual desirability function for the longitudinal expansion index is maximal when that index is 12, which is its target value. By contrast, the individual desirability function for the cross-sectional expansion index is maximal when that index is at its target value of 4. You can also see that the individual desirabilities drop as soon as we are no longer producing on target.

The curves at the bottom of the figure show that a flow rate of 40.6 kg/h, a moisture content of 22.3% and a screw speed of 300 rpm yield an overall desirability of 100%, which means that both of your responses are on target. The middle one of these curves shows that the overall desirability is not very sensitive to the moisture content, as long as it is somewhere between 21.5% and 23.5%.

Of course, I would advise you to carry out a couple of confirmatory runs of your process to verify that the factor settings I suggested do work.

That's about it I think. Perhaps, it is still worth saying that the day-to-day variation in the two responses was about as large as the variation in the residuals. So, you were definitely right to worry about the day-to-day variation.

Let me know if you have further questions.

Half an hour later, Peter finds Maroussa's reply in his mailbox:

Dear Peter:

Thanks for the swift reply. I do have a question though. You wrote me that you used generalized least squares and restricted maximum likelihood estimation to fit the model. Why generalized and not ordinary least squares? And what is restricted maximum likelihood estimation?

Maroussa

Peter responds:

Dear Maroussa:

When you use ordinary least squares or OLS estimation, you assume that the responses for all the runs are uncorrelated. In your study, this assumption is not valid because the responses obtained on any given day are more alike than responses from different days. So, the four responses measured within a given day are correlated. You have to take into account that correlation in order to get the most precise factor-effect estimates and proper significance tests. This is exactly what generalized least squares or GLS estimation does.

To account for that correlation, you first have to quantify it. This is where restricted maximum likelihood or REML estimation comes in.

To understand how REML estimates the correlation between responses collected on the same day, you have to know a bit more about the assumptions behind the model

$$Y_{ij} = \beta_0 + \sum_{k=1}^{3} \beta_k x_{kij} + \sum_{k=1}^{2} \sum_{l=k+1}^{3} \beta_{kl} x_{kij} x_{lij} + \sum_{i=1}^{3} \beta_{kk} x_{kij}^2 + \gamma_i + \varepsilon_{ij} \qquad (7.2)$$

that was in my earlier e-mail (and that we assumed when we generated the design for your study at the conference in Athens). The γ_i and ε_{ij} in the model are random terms, so the model contains two random terms:

1. The term ε_{ij} represents the residual error and captures the run-to-run variation in the responses, just as in the usual response surface models. By assumption, all ε_{ij} are independently normally distributed with zero mean and constant variance σ_ε^2. That variance is called the residual error variance.

2. The second random term in the model is the day or block effect, γ_i. The function of the term γ_i is to model the day-to-day differences between the responses. By assumption, the γ_i are also independently normally distributed with zero mean, but with a different variance, σ_γ^2. The variance σ_γ^2 is called the block variance.

A final assumption about the two random terms in the model is that the random day effects and the residual errors are all independent from each other.

Now, this is where it gets interesting. The interpretation of a small σ_γ^2 value (relative to the residual error variance σ_ε^2) is that the day-to-day variation of the responses is limited, whereas a large value for σ_γ^2 (again, relative to the residual error variance σ_ε^2) indicates a large day-to-day variation. That is easy to see. For example, if σ_γ^2 is small, all γ_i are pretty similar and so are the responses from different days, all other things being equal. Thus, a σ_γ^2 that is small relative to σ_ε^2 indicates that responses from different days do not vary much more than responses from within a given day. An alternative way to put that is to say that a small σ_γ^2 value indicates that responses from the same day have almost nothing in common, or, that the dependence is negligible. To cut a long story short, the magnitude of σ_γ^2 relative to σ_ε^2 measures the dependence in the responses.

The implication of that is that you have to estimate the variances σ_γ^2 and σ_ε^2 to quantify the dependence. Unfortunately, the estimation of variances is always a bit trickier than estimating means or factor effects, and there are a lot of ways to do it. However, REML estimation is a generally applicable procedure that gives you unbiased estimates of variances. Because it is not available in every commercial software package for industrial statisticians, it is still relatively unknown among experimenters in industry. I think that is a shame, since it is a technique that is more than 30 years old already. Also, it is standard practice to use REML in social science and medicine. I don't know how much more detail you want, but in case you want to understand a bit of the conceptual framework behind it, I have attached a short note I wrote to try to demystify REML a little bit (see Attachment 7.1).

Attachment 7.1 Restricted maximum likelihood (REML) estimation.

REML estimation is a general method for estimating variances of random effects in statistical models. It has become the default approach in many disciplines because it guarantees unbiased estimates. The benefits of REML estimation can best be illustrated by considering two situations where a single variance has to be estimated.

1. A well-known estimator of the variance of a population with mean μ and variance σ_ε^2 is the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \tag{7.3}$$

where X_i denotes the ith observation in a random sample of size n. The reason for the division by $n - 1$ rather than by n is to ensure that S^2 is an unbiased estimator of σ_ε^2. The division by $n - 1$ essentially corrects the sample variance for the fact that μ is unknown and needs to be estimated by the sample mean, $\overline{X}$. Now, the sample variance S^2 (involving the division by $n - 1$) can be viewed as the REML estimator of the population variance, σ_ε^2, if the population is assumed to be normally distributed. Dividing by n, which yields the maximum likelihood (ML) estimator, seems like a more intuitive thing to do. However, the ML estimator is biased.

2. The commonly accepted estimator for the residual error variance σ_ε^2 in a regression model with p parameters for analyzing a completely randomized design is the mean squared error (MSE)

$$\text{MSE} = \frac{1}{n-p}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \frac{1}{n-p}\sum_{i=1}^{n} R_i^2, \qquad (7.4)$$

where Y_i and $\hat{Y}_i$ are the observed and predicted responses, respectively, and R_i denotes the ith residual. Here, it is the division by $n-p$ rather than by n which guarantees the unbiasedness of the estimator for the residual error variance. It corrects the MSE for the fact that there are p fixed parameters in the regression model (intercept, main effects, ...) which need to be estimated. Just like the sample variance S^2, you can view the MSE as a REML estimator if the responses in the regression model are normally distributed. Dividing the sum of squared residuals by n yields the ML estimator, which is biased.

These examples illustrate REML's key characteristic: it takes into account that there are location parameters that need to be estimated in addition to variances. Roughly speaking, the only information in the data that REML uses for estimating variances is information that is not used for estimating location parameters, such as means and factor effects. The sample variance in Equation (7.3) takes into account that one unit of information is utilized to estimate the population mean. Similarly, the MSE in Equation (7.4) uses the fact that p information units are absorbed by the estimation of the p regression coefficients. The technical term for these units of information is degrees of freedom.

7.3 Peek into the black box

7.3.1 Model

The model for analyzing data from an experiment run in b blocks of k observations requires an additive block effect for every block. Equation (7.1) gives an example of the model if there are three continuous experimental factors and the interest is in estimating main effects, two-factor interaction effects, and quadratic effects. In general, the model becomes

$$Y_{ij} = \mathbf{f}'(\mathbf{x}_{ij})\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij}, \tag{7.5}$$

where $\mathbf{x}_{ij}$ is a vector that contains the levels of all the experimental factors at the jth run in the ith block, $\mathbf{f}'(\mathbf{x}_{ij})$ is its model expansion, and $\boldsymbol{\beta}$ contains the intercept and all the factor effects that are in the model. For a full quadratic model with three continuous experimental factors,

$$\mathbf{x}_{ij} = \begin{bmatrix} x_{1ij} & x_{2ij} & x_{3ij} \end{bmatrix},$$
$$\mathbf{f}'(\mathbf{x}_{ij}) = \begin{bmatrix} 1 & x_{1ij} & x_{2ij} & x_{3ij} & x_{1ij}x_{2ij} & x_{1ij}x_{3ij} & x_{2ij}x_{3ij} & x_{1ij}^2 & x_{2ij}^2 & x_{3ij}^2 \end{bmatrix},$$

and

$$\boldsymbol{\beta}' = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 & \beta_{12} & \beta_{13} & \beta_{23} & \beta_{11} & \beta_{22} & \beta_{33} \end{bmatrix}.$$

As in the model in Equation (7.1), γ_i represents the ith block effect and ε_{ij} is the residual error associated with the jth run on day i. If successful, including block effects in a model substantially reduces the unexplained variation in the responses. This results in higher power for the significance tests of the factor effects. So, including block effects in the model often results in higher chances for detecting active effects.

We assume that the block effects and the residual errors are random effects that are all independent and normally distributed with zero mean, that the block effects have variance σ_γ^2, and that the residual errors have variance σ_ε^2. The variance of a response Y_{ij} is

$$\text{var}(Y_{ij}) = \text{var}(\gamma_i + \varepsilon_{ij}) = \text{var}(\gamma_i) + \text{var}(\varepsilon_{ij}) = \sigma_\gamma^2 + \sigma_\varepsilon^2,$$

whereas the covariance between each pair of responses Y_{ij} and $Y_{ij'}$ from the same block i is

$$\text{cov}(Y_{ij}, Y_{ij'}) = \text{cov}(\gamma_i + \varepsilon_{ij}, \gamma_i + \varepsilon_{ij'}) = \text{cov}(\gamma_i, \gamma_i) = \text{var}(\gamma_i) = \sigma_\gamma^2,$$

and the covariance between each pair of responses Y_{ij} and $Y_{i'j'}$ from different blocks i and i' equals

$$\text{cov}(Y_{ij}, Y_{i'j'}) = \text{cov}(\gamma_i + \varepsilon_{ij}, \gamma_{i'} + \varepsilon_{i'j'}) = 0.$$

This is due to the independence of all the block effects and the residual errors, which implies that every pairwise covariance between two different random effects is zero. The correlation between a pair of responses from a given block i is

$$\text{corr}(Y_{ij}, Y_{ij'}) = \frac{\text{cov}(Y_{ij}, Y_{ij'})}{\sqrt{\text{var}(Y_{ij})}\sqrt{\text{var}(Y_{ij'})}} = \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\varepsilon^2}.$$

This within-block correlation coefficient increases with σ_γ^2. So, the larger the differences between the blocks (measured by σ_γ^2), the larger the similarity between every pair of responses from a given block. The dependence between responses from the same block, $Y_{i1}, \ldots, Y_{ik}$, is given by the symmetric $k \times k$ matrix

$$\mathbf{\Lambda} = \begin{bmatrix} \sigma_\gamma^2 + \sigma_\varepsilon^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 \\ \sigma_\gamma^2 & \sigma_\gamma^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \cdots & \sigma_\gamma^2 + \sigma_\varepsilon^2 \end{bmatrix}, \tag{7.6}$$

which is the variance-covariance matrix of all the responses in block i. The jth diagonal element of that matrix is the variance of Y_{ij}, and the off-diagonal element in row j and column j' is the covariance between Y_{ij} and $Y_{ij'}$.

A key assumption of the model in Equation (7.5) is that the block effects are additive, which implies that each block has a random intercept. Under this assumption, the factor effects do not vary from block to block. Obviously, this is a strong assumption, but it has been confirmed in many empirical studies. It is possible to allow the factor effects to vary from one block to another by including additional random effects in the model, but this falls outside the scope of this book.

Note also that we focus on scenarios where the number of runs is the same for every block. While we believe that this scenario is the most common one, the model can easily be adapted to handle experiments with blocks of different sizes.

The model in Equation (7.5) is called a mixed model, because it involves two sorts of effects. The factor effects are fixed effects, whereas the block effects and the residual errors are random effects. The distinction between the two types of effects is that the latter are assumed to be random variables, whereas the former are not. An interesting property of the mixed model for analyzing data from blocked experiments is that quantifying the differences between the blocks is done by estimating σ_γ^2. So, no matter how many blocks an experiment involves, it is only necessary to estimate one parameter to account for the block effects. An alternative model specification would treat the block effects as fixed effects, rather than random ones. However, the model specification involving fixed block effects requires the estimation of $b - 1$ parameters to quantify the differences between the blocks, and therefore necessitates a larger number of experimental runs. This is a major advantage of the mixed model. The next chapter deals explicitly with this issue.

7.3.2 Generalized least squares estimation

In Section 7.3.1, we showed that responses in the model in Equation (7.5) from the same block are correlated. This violates the independence assumption of the ordinary least squares or OLS estimator. While that estimator is still unbiased, it is better to use generalized least squares or GLS estimates, which account for the correlation and are generally more precise. The GLS estimates are calculated from

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\boldsymbol{Y}. \tag{7.7}$$

It is different from the OLS estimator,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{Y},$$

in that it is a function of the matrix $\mathbf{V}$ that describes the covariance between the responses.

To derive the exact nature of $\mathbf{V}$, we rewrite the model for analyzing data from blocked experiments in Equation (7.5) as

$$\boldsymbol{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{7.8}$$

where

$$\boldsymbol{Y} = \begin{bmatrix} Y_{11} & \cdots & Y_{1k} & Y_{21} & \cdots & Y_{2k} & \cdots & Y_{b1} & \cdots & Y_{bk} \end{bmatrix}',$$

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_b \end{bmatrix} = \begin{bmatrix} \mathbf{f}'(\mathbf{x}_{11}) \\ \vdots \\ \mathbf{f}'(\mathbf{x}_{1k}) \\ \mathbf{f}'(\mathbf{x}_{21}) \\ \vdots \\ \mathbf{f}'(\mathbf{x}_{2k}) \\ \vdots \\ \mathbf{f}'(\mathbf{x}_{b1}) \\ \vdots \\ \mathbf{f}'(\mathbf{x}_{bk}) \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_b \end{bmatrix}',$$

and

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_{11} & \cdots & \varepsilon_{1k} & \varepsilon_{21} & \cdots & \varepsilon_{2k} & \cdots & \varepsilon_{b1} & \cdots & \varepsilon_{bk} \end{bmatrix}'.$$

Note that the elements of $\boldsymbol{Y}$ and the rows of $\mathbf{X}$ are arranged per block, and that the model matrix is composed of one submatrix $\mathbf{X}_i$ for every block. The matrix $\mathbf{Z}$ is a matrix that has a one in the ith row and jth column if the ith run of the experiment belongs to the jth block, and a zero otherwise. For example, the $\mathbf{Z}$ matrix has ones in its first column for the first k rows, indicating that the first k runs belong to the first block. Because we assume that all the random effects are independent, the random-effects vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ have variance–covariance matrices $\text{var}(\boldsymbol{\gamma}) = \sigma_\gamma^2\mathbf{I}_b$

and $\text{var}(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{I}_n$, respectively, where $\mathbf{I}_b$ and $\mathbf{I}_n$ are identity matrices of dimensions b and n, respectively.

The matrix $\mathbf{V}$ is the variance–covariance matrix of the response vector:

$$\begin{aligned}
\mathbf{V} &= \text{var}(\boldsymbol{Y}),\\
&= \text{var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}),\\
&= \text{var}(\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}),\\
&= \text{var}(\mathbf{Z}\boldsymbol{\gamma}) + \text{var}(\boldsymbol{\varepsilon}),\\
&= \mathbf{Z}[\text{var}(\boldsymbol{\gamma})]\mathbf{Z}' + \text{var}(\boldsymbol{\varepsilon}),\\
&= \mathbf{Z}[\sigma_\gamma^2 \mathbf{I}_b]\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n,\\
&= \sigma_\gamma^2 \mathbf{Z}\mathbf{Z}' + \sigma_\varepsilon^2 \mathbf{I}_n,
\end{aligned}$$

which is an $n \times n$ block diagonal matrix of the form

$$\mathbf{V} = \begin{bmatrix} \boldsymbol{\Lambda} & \mathbf{0}_{k\times k} & \cdots & \mathbf{0}_{k\times k} \\ \mathbf{0}_{k\times k} & \boldsymbol{\Lambda} & \cdots & \mathbf{0}_{k\times k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{k\times k} & \mathbf{0}_{k\times k} & \cdots & \boldsymbol{\Lambda} \end{bmatrix},$$

where $\boldsymbol{\Lambda}$ is the variance-covariance matrix for all the responses within a given block given in Equation (7.6) and $\mathbf{0}_{k\times k}$ is a $k \times k$ matrix of zeros. The interpretation of the elements of $\mathbf{V}$ is similar to that for the elements of $\boldsymbol{\Lambda}$. The diagonal elements of $\mathbf{V}$ are the variances of the responses Y_{ij}, and the off-diagonal elements are covariances between pairs of responses. The nonzero off-diagonal elements of $\mathbf{V}$ all correspond to pairs of responses from within a given block. All zero elements of $\mathbf{V}$ correspond to pairs of runs from two different blocks.

The GLS estimator in Equation (7.7) is a function of the matrix $\mathbf{V}$, and this complicates the estimation of the factor effects. This is because, even though the structure of $\mathbf{V}$ is known, we do not know the exact value of its nonzero elements. Therefore, these nonzero elements, which are functions of the block variance σ_γ^2 and the residual error variance σ_ε^2, have to be estimated from the data. So, to use the GLS estimator, we have to estimate σ_γ^2 and σ_ε^2, and insert the estimates into the covariance matrices $\boldsymbol{\Lambda}$ and $\mathbf{V}$. Denoting the estimators for σ_γ^2 and σ_ε^2 by $\hat{\sigma}_\gamma^2$ and $\hat{\sigma}_\varepsilon^2$, respectively, the estimated covariance matrices can be written as

$$\hat{\boldsymbol{\Lambda}} = \begin{bmatrix} \hat{\sigma}_\gamma^2 + \hat{\sigma}_\varepsilon^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 + \hat{\sigma}_\varepsilon^2 & \cdots & \hat{\sigma}_\gamma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_\gamma^2 & \hat{\sigma}_\gamma^2 & \cdots & \hat{\sigma}_\gamma^2 + \hat{\sigma}_\varepsilon^2 \end{bmatrix} \tag{7.9}$$

and

$$\widehat{\mathbf{V}} = \begin{bmatrix} \hat{\mathbf{\Lambda}} & \mathbf{0}_{k\times k} & \cdots & \mathbf{0}_{k\times k} \\ \mathbf{0}_{k\times k} & \hat{\mathbf{\Lambda}} & \cdots & \mathbf{0}_{k\times k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{k\times k} & \mathbf{0}_{k\times k} & \cdots & \hat{\mathbf{\Lambda}} \end{bmatrix}.$$

The GLS estimator of the factor effects then becomes

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\boldsymbol{Y}. \tag{7.10}$$

That estimator is an unbiased estimator of $\boldsymbol{\beta}$, and is often named the feasible GLS estimator.

7.3.3 Estimation of variance components

There are different ways to estimate the variances σ_γ^2 and σ_ε^2 in an unbiased fashion, including an analysis of variance method called Henderson's method or Yates's method and REML (restricted or residual maximum likelihood) estimation. Henderson's method is computationally simple, and does not require the assumption of normality. It does, however, yield less precise estimates than REML if the fitted model is correctly specified. While REML estimation assumes that all the random effects are normally distributed, simulation results have indicated that it still produces unbiased estimates if that assumption is violated. Strong arguments in favor of REML estimates are that they have minimum variance in case the design is balanced and that, unlike other types of estimates, they are calculable for unbalanced designs too. For these reasons, and due to its availability in commercial software, REML estimation of variances in mixed models is the default method in many disciplines. For two familiar REML type estimators, we refer to Attachment 7.1 on page 150–151.

The REML estimates for σ_γ^2 and σ_ε^2 are obtained by maximizing the following restricted log likelihood function:

$$l_R = -\frac{1}{2}\log|\mathbf{V}| - \frac{1}{2}\log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}\mathbf{r}'\mathbf{V}^{-1}\mathbf{r} - \frac{n-p}{2}\log(2\pi),$$

where $\mathbf{r} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and p is the number of parameters in $\boldsymbol{\beta}$. Like any other method, except for Bayesian methods, REML may provide a negative estimate of σ_γ^2 even in situations where the block-to-block variance is expected to be well above zero. The chances for getting a negative estimate for σ_γ^2 are larger when the number of blocks is small, the design is not orthogonally blocked and σ_γ^2 is small compared to σ_ε^2. In the event of such a negative estimate, we recommend setting

σ_γ^2 to a positive value when computing the GLS estimator to see how sensitive the factor-effect estimates, their standard errors and the p values are to its value, or to use a Bayesian analysis method.

7.3.4 Significance tests

To perform significance tests for each of the factor effects, three pieces of information are needed for each of the parameters in $\boldsymbol{\beta}$: the parameter estimate, its standard error, and the degrees of freedom for the t distribution to compute a p value. The parameter estimates are given by the feasible GLS estimator in Equation (7.10).

Finding the standard errors is more involved. If the variances σ_γ^2 and σ_ε^2 were known, then the standard errors would be the square roots of the diagonal elements of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$. However, σ_γ^2 and σ_ε^2 are unknown, and the logical thing to do is to substitute σ_γ^2 and σ_ε^2 by their estimates in $\mathbf{V}$. In this approach, the standard errors are the square roots of the diagonal elements of

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}. \tag{7.11}$$

In general, however, the standard errors obtained in this way underestimate the true standard errors a little bit. Slightly better standard errors can be obtained by using the diagonal elements from an adjusted covariance matrix

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1} + \boldsymbol{\Omega} \tag{7.12}$$

instead of $(\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}$. The derivation of the adjustment, $\boldsymbol{\Omega}$, is too complex to give here, but it can be found in Kenward and Roger (1997), to whom the adjustment of the standard errors is due. The exact adjustment depends on the design, and, for designs which are orthogonally blocked or nearly orthogonally blocked (such as optimal designs), the adjustment is small (and sometimes even zero for some or all of the factor effects). Kenward and Roger (1997) also describe a method for determining the degrees of freedom for hypothesis tests based on the adjusted standard errors. The main software packages have already implemented the adjusted standard errors and degrees of freedom.

7.3.5 Optimal design of blocked experiments

To find an optimal design for a blocked experiment, we have to solve two decision problems simultaneously. First, we have to determine factor levels for each run, and, second, we have to arrange the selected factor-level combinations in b blocks of size k. The challenge is to do this in such a way that we obtain the most precise estimates possible. Therefore, we seek factor-level combinations and arrange them in blocks so that the covariance matrix of the GLS estimator, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, is as small as possible. The way in which we do this is by minimizing the determinant of the covariance matrix $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$, which yields the same result as maximizing the determinant of the information matrix $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$. The block design that minimizes the determinant of the GLS estimator's covariance matrix or maximizes the determinant of the information matrix is the D-optimal block design. This results both in small variances of the parameter estimates and in small covariances between the parameter estimates. To find a D-optimal block design, we utilize the coordinate-exchange algorithm outlined in Section 2.3.9, but we use $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$ instead of $|\mathbf{X}'\mathbf{X}|$ as the objective function to maximize.

A technical problem with finding the D-optimal design for a blocked experiment is that the matrix $\mathbf{V}$, and therefore also the D-optimality criterion $|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|$, depends on the unknown variances σ_γ^2 and σ_ε^2. It turns out that the D-optimal block design does not depend on the absolute magnitude of these two variances, but only on their relative magnitude. As a result, software to generate optimal block designs requires input on the relative magnitude of σ_γ^2 and σ_ε^2. Our experience suggests that the block-to-block variance, σ_γ^2, is often substantially larger than the run-to-run variance, σ_ε^2. So, we suggest specifying that the variance ratio $\sigma_\gamma^2/\sigma_\varepsilon^2$ is at least one when generating a D-optimal block design. For the purpose of generating an excellent design, an educated guess of the variance ratio is good enough because the D-optimal design is not sensitive to the specified value. So, a design that is optimal for one variance ratio is also optimal for a broad range of variance ratios smaller and larger than the specified one. Moreover, in the rare cases where different variance ratios lead to different designs, the quality of these designs is almost indistinguishable. Goos (2002) recommends using a variance ratio of one for finding optimal designs for blocked experiments in the absence of detailed a priori information about it.

7.3.6 Orthogonal blocking

Textbooks on experimental design and response surface methodology emphasize the importance of orthogonal blocking. Orthogonally blocked designs guarantee factor-effect estimates that are independent of the estimates of the block effects. The general condition for orthogonal blocking is that

$$\frac{1}{k}\sum_{j=1}^{k}\mathbf{f}(\mathbf{x}_{ij}) = \frac{1}{k}\mathbf{X}_i'\mathbf{1}_k = \frac{1}{n}\mathbf{X}'\mathbf{1}_n = \frac{1}{n}\sum_{i=1}^{b}\sum_{j=1}^{k}\mathbf{f}(\mathbf{x}_{ij}) \tag{7.13}$$

for each of the b blocks. The interpretation of this condition is that, in an orthogonally blocked design, every $\mathbf{X}_i$ has the same average row. The condition for orthogonal blocking extends to designs with unequal numbers of runs in their blocks.

As an example, consider the orthogonally blocked central composite design in Table 7.1, meant for estimating the model in Equation (7.1). The design consists of three blocks, so the model matrix for that design consists of three submatrices, $\mathbf{X}_1$,

$\mathbf{X}_2$, and $\mathbf{X}_3$, one for each block:

$$
\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} = \left[\begin{array}{cccccccccc}
1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 \\
1 & 1 & 1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\hline
1 & -1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 & 0 & 0 \\
1 & 1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 & 0 & 0 \\
1 & 0 & -1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 & 0 \\
1 & 0 & 1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 & 0 \\
1 & 0 & 0 & -1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 \\
1 & 0 & 0 & 1.633 & 0 & 0 & 0 & 0 & 0 & 2.667 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
$$

For each of these three submatrices, the average row is

$$
\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2/3 & 2/3 & 2/3 \end{bmatrix},
$$

where the first element corresponds to the intercept and the next three sets of three elements correspond to the main effects, the two-factor interaction effects, and the quadratic effects, respectively. The D-optimal design in Table 7.5 is not orthogonally blocked because the row averages of the submatrices $\mathbf{X}_i$ of the model matrix $\mathbf{X}$ are unequal. Table 7.10 shows this.

You can quantify the extent to which a design is orthogonally blocked in various ways. One approach is to compute an efficiency factor for each of the effects in the

Table 7.10 Average rows of the matrices $\mathbf{X}_i$ for each of the seven blocks of the D-optimal design in Table 7.5. The column for the intercept is not displayed.

	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
1	0.25	−0.25	−0.25	0.25	0.25	−0.25	0.75	0.75	0.75
2	0.25	0.25	0.25	0.00	0.00	0.00	0.75	0.75	0.75
3	0.25	0.25	−0.25	0.00	0.00	0.00	0.75	0.75	0.75
4	−0.25	−0.25	0.25	−0.25	0.25	0.25	0.75	0.75	0.75
5	−0.25	0.25	0.25	0.00	0.00	0.00	0.75	0.75	0.75
6	0.00	0.00	0.00	0.00	0.00	0.00	1.00	1.00	1.00
7	−0.25	0.25	−0.25	0.00	0.00	0.00	0.75	0.75	0.75

model. The efficiency factor is the ratio of the variance of the factor-effect estimate in the absence of block effects to the variance in the presence of block effects. If a design is orthogonally blocked, these two variances are identical because, in that case, the factor-effect estimates are independent of the block effects' estimates. This results in an efficiency factor of 100%. If a design is not orthogonally blocked, then the variance of some of the factor-effect estimates is inflated to some extent. This results in efficiency factors of less than 100%. The variances in the presence and absence of block effects required to compute the efficiency factors are the diagonal elements of the matrices $([\mathbf{X}\ \mathbf{Z}]'[\mathbf{X}\ \mathbf{Z}])^{-1}$ and $(\mathbf{X}'\mathbf{X})^{-1}$, respectively. The efficiency factors do not depend on the variances σ_γ^2 and σ_ε^2.

The inverse of the efficiency factors gives the variance inflation factors or VIFs due to the blocking. For the pastry dough experiment, the smallest efficiency factor is 93.4%, so the largest VIF is $1/0.934 = 1.0705$. This was shown in Table 7.6.

7.3.7 Optimal versus orthogonal blocking

The complexity of many experimental design problems is such that it is often difficult, if not impossible, to find an orthogonally blocked design. As a result, there is often no other possibility than to generate an optimal experimental design, that is not orthogonally blocked, for a practical experiment that has to be run in blocks.

However, there also exist scenarios where it is possible to construct an orthogonally blocked design, and where that design is in fact D-optimal. So, there is not necessarily a conflict between orthogonality and optimality of a design. As a matter of fact, Goos and Vandebroek (2001) provide theoretical evidence that D-optimal designs tend to be orthogonally blocked.

The reverse, however, is not generally true. It may be possible to construct orthogonally blocked designs for certain practical problems, which are not even close to being optimal. Goos and Donev (2006b) provide an example of an optimal block design which yields variances for the factor-effect estimates which are three times smaller than those produced by an orthogonally blocked alternative described in the literature.

It is our view that whether or not a design is orthogonally blocked is of secondary importance. What is of primary importance is that the factor-effect estimates can be estimated precisely, so that powerful statistical tests and precise predictions are possible. The results in Goos and Vandebroek (2001) indicate that a design that produces precise factor-effect estimates is automatically at least nearly orthogonally blocked.

7.4 Background reading

More details about the optimal design of experiments in blocks can be found in Chapters 4 and 5 of Goos (2002). These chapters review much of the literature on the subject, provide additional examples of optimal block designs, and discuss in detail the design of an optometry experiment. The optometry experiment involves one continuous factor with a quadratic effect. The blocks in the experiment are test subjects and have two observations each. Further discussion about the optimality and orthogonality of block designs is given in Goos (2006).

More technical details and discussion about the analysis of data from blocked experiments are given in Khuri (1992), Gilmour and Trinca (2000), and Goos and Vandebroek (2004). The REML method, which is recommended by Gilmour and Trinca (2000) for analyzing data from blocked experiments, was originally proposed by Patterson and Thompson (1971). The unbiasedness and variance of the feasible GLS estimator in Equation (7.10) are discussed in Kackar and Harville (1981, 1984), Harville and Jeske (1992), and Kenward and Roger (1997). Gilmour and Goos (2009) show how to use a Bayesian approach to overcome the problem of a negative estimate for σ_γ^2 for data from a split-plot experiment. Khuri (1996) shows how to handle data from blocked experiments when the block effects are not additive.

This chapter presented a case study with only one blocking factor, the day on which an experimental run was carried out. Gilmour and Trinca (2003) and Goos and Donev (2006a) discuss extensions to situations where there is more than one blocking factor. Goos and Donev (2006b) demonstrate the benefits of optimal design for setting up mixture experiments in blocks.

The case study in this chapter features two different responses. For the simultaneous optimization of these responses, we used a desirability function. For more details about desirability functions, we refer to Harrington (1965) and Derringer and Suich (1980).

7.5 Summary

Practical considerations often cause experimental runs to appear in groups. Common causes for this are that the experiment involves several days of work or that several batches of raw material are used in an experiment. Experiments in the semi-conductor industry, for example, often require the use of different silicon wafers. Other examples of experiments where the runs are grouped are agricultural experiments where multiple fields are used and medical, marketing, or taste experiments where each of different subjects tests and rates several factor-level combinations.

The groups of runs are called blocks, and the factor that causes the grouping is the blocking factor. The block-to-block differences often account for a large share of the overall variability in the responses, and necessitate a careful planning of the experiment and subsequent data analysis. Therefore, we recommend spending substantial effort on the identification of sources of variation such as blocking factors when planning an experimental study. This is especially important when the block-to-block variability is comparable to or larger than the run-to-run variability.

In this chapter, we showed how a mixed model, involving random block effects, can be used as a starting point for generating an optimal block design and for the analysis of the data from that design. In the next chapter, we discuss the design of another blocked experiment, and highlight the benefits of the random block effects model compared to the fixed block effects model.