Using Raw Material Measurements in Robust Process Optimization ⋆

Bjørn-Helge Mevik [1], Ellen Mosleth Færgestad,
Marit Risberg Ellekjær and Tormod Næs

MATFORSK, Osloveien 1, N-1430 Ås, Norway.

Abstract

Unwanted variation due to variable raw material quality is often a problem in production processes. Robust process optimization seeks to reduce the effects of such variation by identifying settings of the adjustable factors that makes the process less sensitive to the variations. This paper develops a unified framework for studying and developing robust process optimization and process control techniques. We divide the factors of the process into groups based on characterizations of their properties. We also develop a robust process optimization technique for batch-wise processes, called batch-wise robust process optimization, which utilizes all available measurements of raw material qualities at the start of each production batch. The technique achieves a reduction of variability due to variation in raw material qualities, compared to ordinary robust process optimization. Two examples taken from baking of hearth bread illustrate the technique.

Key words: Raw Material Measurements; Batch-Wise Robust Process Optimization; Taguchi Method; Process Control; Robust Parameter Design

1 Introduction

Industrial production processes frequently suffer from unwanted variation in raw material quality, process settings and environmental factors. Such variation may result in undesirable variation in the quality of the end product.

⋆ This is a preprint of an article published in *Chemometrics and Intelligent Laboratory Systems* 55 (2001) 133–145

[1] Corresponding author. Tel: +47 64970338; Fax: +47 64970333; E-mail: Bjorn-Helge.Mevik@matforsk.no

The present paper concerns some methodological aspects related to this problem. The type of process we have in mind is a batch-wise production process where there is batch-wise variation in the quality of the raw material, but negligible variation within each batch. The quality of the raw material can be characterized by some rapid measurement method, and is available at the start of production of each batch.

Two common approaches to cope with the consequences of such variation are:

Use statistical process control (SPC) or engineering process control (EPC) techniques to monitor and adjust the factors. SPC is used to monitor the process using statistical methods in order to give operators diagnostic displays of the state of the process. EPC uses feed-forward/feedback regulation techniques to control the process. SPC and EPC have many successful applications [1]. The methods share the possible drawback that continuously monitoring or adjusting the process may be cumbersome and expensive.

Use robust process optimization to make the production as insensitive as possible to variations in the factors. Robust process optimization is an off-line quality control method, and is known under several names, sometimes with slightly different definitions [2–13]. Robust process optimization tries to adjust the process *before* the production starts, to minimize the transmitted variation. It offers the benefit of reduced need for continuously monitoring and adjusting the process, without having to adjust factors that are expensive to control. However, it often requires more experiments in front of the analysis than SPC or EPC, which in some situations can be expensive. Also, sometimes, the reduction in transmitted variation achieved is not big enough, and SPC or EPC might be needed to further control the process variation during production.

The aim of this paper is twofold. One goal is to develop a general, flexible and unified framework in which one can study and develop process optimization and process control methods. We achieve this goal by sorting the different factors that influence the quality of the product into groups according to their inherent properties.

The other goal is to present a combination of the control and robustness strategies described above. This will be called *batch-wise robust process optimization*. The approach combines the ideas of robust process optimization and feed-forward regulation. All possible information available at the start of a production batch is used to identify process settings which make the end product as insensitive as possible to other sources of noise during production of the batch.

The theory of the paper will be illustrated by two examples of optimization of a baking process. The examples are taken from an ongoing effort to improve bread quality based on knowledge of the raw material in combination with process settings such as mixing time and proofing time.

1.1 Notation

Scalars will be denoted by lower case letters, and vectors and vector-valued functions will be denoted by bold lower case letters. We denote expectation by E, expectation with respect to the joint distribution of x and y by $\mathsf{E}_{x,y}$, and the conditional expectation of x given a by $\mathsf{E}[x \mid a]$.

2 Model and Example

2.1 The Model

The present approach to robust optimization will be based on an estimated model of the relationship between the input variables and the end product quality. We assume that the process can be described by a mathematical model

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_q)$ is a vector of one or more response variables, $\boldsymbol{x}$ is a vector of model variables, $\boldsymbol{f}$ is a vector-valued function, and $\boldsymbol{\varepsilon}$ is the error term. The error is assumed to be random with zero expectation, independent of $\boldsymbol{x}$ and having constant variance.

The response variables represent the quality characteristics of the product or process. Typical characteristics in the example to be considered next are color or loaf volume of baked bread. Each response variable y_i has a target value t_i, representing the preferred values of the quality characteristics, defined either by consumers or experienced bakers. They can be finite or infinite. The model variables $\boldsymbol{x}$ represent factors of the process that influence the quality characteristics of the product. These factors are typically raw material quality variables, process settings and environmental factors such as temperature and humidity.

In most cases, the function $\boldsymbol{f}$ is unknown and has to be estimated from data. In practice, one will typically set up an experimental design based on an established strategy appropriate for the situation [14,15]. The data from this design are most frequently fitted to an empirical model. For the purpose of the methods to be developed in this paper, we assume that the function is already estimated or known. In the example below, a combination of a factorial and mixtures design is used to provide the data. A regular second degree model is used.

Note that factors may be controllable and possible to incorporate in a design during experimentation, but totally uncontrollable during a regular operation. Raw material quality is a typical example of this. It may be possible to set up a reasonable design in predefined quality classes, but during regular operation one has to take what is available. The focus in this paper is on methodology and ideas related to regular operation.

2.2 Example

Nomenclature and data from baking is used to illustrate the ideas and analysis techniques. In baking processes there are both a large number of raw material properties and several process settings that are important and can be varied.

Typical process steps that can influence the quality of the loaves are mixing time, proofing time and baking time. The temperatures in the different process steps are also important. All these variables can in principle be varied. Some of them can be controlled easily, while others are uncontrollable.

This example is used to discuss the different categories of factors that are possible in practice, followed by a real baking experiment to illustrate the new methodology.

3 Classification of the Factors

For the purpose of this paper, we will think of a process factor as a stochastic variable following a distribution during regular operation of the process. The distribution is governed by a vector of distribution parameters, which can be viewed as a point in a parameter space that includes all valid values of the parameters. During normal production one may be able to adjust some of these parameters while others are fixed. Baking temperature is a typical process factor. If the oven temperature varies, it is a stochastic variable. For simplicity, assume that it is normally distributed. The distribution parameters are then the mean temperature and the variance. The mean is usually adjustable, while we would regard the variance as a fixed parameter.

The factors can be categorized in different ways. We have chosen a classification based on the properties of the factors during normal production.

A factor will be called *adjustable* if we can adjust *some* of its distribution parameters and *unadjustable* if *all* distribution parameters are fixed. The baking temperature is then adjustable, because we can adjust its mean. Note that according to this definition we do not require *total* control over a factor to call it adjustable.

Similarly, a factor will be called *constant* if it has zero variance, and *variable* if it has positive variance. Modeling a constant factor as a stochastic variable with zero variance allows us to treat constant and variable factors similarly. The baking temperature above is variable.

Another distinction is whether the factors are *observable* or *unobservable*. A factor is observable if its measured value can be used in the optimization or control procedure. For batch processes, a factor has to be constant during production of the batch, and measurable at the time the production of the batch starts, to be considered observable.

	Observable		Unobservable	
	Constant	Variable	Constant	Variable
Adjustable	Design parameter (A)	Design parameter with observable variation (B)	*Impossible*	Design parameter with noise (C)
Unadjustable	*Constant*	Observable uncontrollable factor (D)	*Unknown constant*	Noise (E)

Table 1
Grouping of factors. The entries are descriptions of the group, or traditional names of typical factors belonging to the group. The letters refer to factors in the example of Section 3.1.

Together, these criteria split the factors into eight groups, as displayed in Table 1.

Three of the factor groups are special. Firstly, unobservable, adjustable, constant factors do not exist. Secondly, unadjustable, constant factors, whether observable or unobservable, are simply constants and will enter the process model (1) as such. These groups are not interesting for the purposes of process optimization or control and are included here for symmetry only.

3.1 A Constructed Example to Illustrate the Groups of Factors

To illustrate the groups, we imagine a baking process modeled by

$$\boldsymbol{y} = \boldsymbol{f}(A, B, C, D, E) + \boldsymbol{\varepsilon}, \qquad (2)$$

for some function $\boldsymbol{f}$. Here $\boldsymbol{y}$ is a vector of responses related to the quality of the bread and $\boldsymbol{\varepsilon}$ is model error. There are two process variables; room temperature B and proofing temperature C; and three raw material properties A, D and E, one for each of three different raw materials; R_A, R_D and R_E, respectively. All other factors are considered constant. The letters A, to E correspond to the letters in Table 1.

We assume that the quality A of raw material R_A, for instance fat, can be completely controlled. It can be set to a value μ_A, and does not vary. Regarded as a stochastic variable, it has a constant distribution with mean μ_A. Thus, A is an observable, adjustable and constant factor. Traditionally these factors are called *design parameters*, *design factors* or *product parameters* [2,3,16,17].

Room temparature B can be set to a nominal value μ_B, but varies around its setting. We assume that it can be measured, so it is observable, adjustable and variable. Since the variation of these factors is observable, we may use it in an optimization or control procedure. However, because the value in general cannot be measured until after the factor has been given a nominal value, it is not available until after the optimization. To use such a value, one might perform a second optimization, treating the observed value as the fixed value of the factor.

Proofing temperature C can be set to a nominal value μ_C, and varies around its setting during production. We assume that it cannot be measured during normal production. The proofing temperature belongs to the group of unobservable, adjustable, variable factors. These factors are often called *design factors with noise* [4,18]. Taguchi [6,7] calls the variation of these factors *inner noise.*

The moisture content of the raw material R_D, for instance flour, cannot be controlled, and we assume it varies according to a fixed distribution. We assume we can measure the value of D. It follows that D is an observable, unadjustable and variable factor. We shall see that for batch-wise robust process optimization, we do not need the distribution of D, only its measured value. Pledger [19] uses the term *observable uncontrollable factor* for a factor that is measurable at all times, and uses the measurements to continuously adjust the process.

Assume that the raw material quality E, for instance yeast viability follows a fixed distribution and cannot be controlled or measured during normal production. Both the raw material quality E and the model error $\boldsymbol{\varepsilon}$ are unobservable, unadjustable and variable factors. In the terminology of Taguchi, E is a *noise factor*, while $\boldsymbol{\varepsilon}$ is *pure noise.*

4 Batch-Wise Robust Process Optimization

In the terminology of the previous section, robust process optimization now amounts to specifying a criterion for robustness and minimize it with respect to the adjustable distribution parameters of the factors. For a single-response system, it is natural to specify a loss function of y and use its expectation over the variable factors as the optimization criterion. One of the most widely used loss functions is the quadratic loss. With a finite target value, the quadratic loss function is $L(y) = c(y - t)^2$, for some constant c. The quadratic loss function is simple, has convenient mathematical properties and is easy to interpret [3], but other possibilities also exist.

Loss functions are not the only robustness criteria. For instance, Taguchi [6,7,17] uses signal-to-noise ratios, and Box and Jones [10] note that $\mathsf{E}\, L(y) = (\mathsf{E}[y] - t)^2 + \mathrm{Var}(y)$, and propose a generalization given by $w(\mathsf{E}[y] - t)^2 + (1 - w)\,\mathrm{Var}(y)$, for $w \in [0, 1]$.

For batch-wise robust process optimization we first compute the loss function conditional on the parameters that can be controlled. This is done at the start of each batch of the production, and the values of all observable, unadjustable factors are used. The loss function given the values of the controllable parameters, is then a stochastic variable. The expectation of this variable, the *expected loss*, is therefore a function of these given parameters. Finally, the expected loss function is minimized, giving the optimal nominal values of the adjustable factors.

The idea behind this approach is that we measure all the noise we can observe, use this information when adjusting the parameters we can control, and leave the remaining unobservable noise untreated. In this way, we use all available information when adjusting the process to compensate for the remaining noise. This is a kind of compromise between the two possibilities mentioned in the introduction.

To illustrate the procedure, we apply it to the example in the previous section. For each batch, we first measure the raw material quality D (moisture content of the flour), as this is an observable, unadjustable factor. We denote the measured value D_m. We enter this value into the process model (2). The expected loss is

$$\mathsf{E}_{A,B,C,E}\left[L(A, B, C, D, E) \mid \mu_A, \mu_B, \mu_C, D = D_\mathrm{m}\right], \tag{3}$$

where the μ's are the mean parameters of the factors. Since factor A is adjustable and constant, the expected loss can be written as

$$\mathsf{E}_{B,C,E}[L(A, B, C, D, E) \mid A = \mu_A, \mu_B, \mu_C, D = D_\mathrm{m}]. \tag{4}$$

We then minimize this function with respect to μ_A, μ_B, and μ_C, and arrive at the optimal parameter settings $(\mu_A^*, \mu_B^*, \mu_C^*)$. This gives the minimal expected loss

$$g(\mu_A^*, \mu_B^*, \mu_C^*) = \mathsf{E}_{B,C,E}[L(A, B, C, D, E) \mid A = \mu_A^*, \mu_B^*, \mu_C^*, D = D_\mathrm{m}]. \quad (5)$$

4.1 Multi-Response Systems

Systems with more than one response are also in this case more complex than single-response systems. It is possible that the optimal point of one criterion turns out to be optimal or nearly optimal for some of the other criteria as well, but this is seldom the case.

One common approach is to specify loss functions for each response variable and combine them into a total loss. The simplest form is a weighted sum of the losses, which we use here. One then uses the expected total loss as the minimization criterion. It is often useful to apply some sort of standardization of the responses or loss functions, in order to make it easier to specify or interpret the weights. We have chosen to standardize the loss functions by dividing them by their standard deviation. This has the benefit that the quadratic loss functions become invariant to the scale of the response.

The minimization of the robustness criterion can optionally be subject to restrictions, either on the model variables or on the responses. Examples include keeping the model variables inside a certain region to ensure operational solutions, or requiring that a response should be unbiased, i.e., $\mathsf{E}\, y_i = t_i$.

5 Examples

Both our examples are concerned with the baking of hearth bread. The situation in these examples is a baking process where the flour arrives in batches. Between each batch there is substantial variation in protein content, but only negligible variation within the batches. The protein content of flour can be assessed by Near Infrared Reflectance (NIR) measurements [20], and will therefore be available at production start for each batch. It is therefore an observable, unadjustable and variable factor.

Two other factors of interest are the mixing and proofing time of the dough. The mixing time can be completely controlled, and that proofing time can be adjusted, but varies randomly around its setting. The mixing time is therefore an observable, adjustable and constant factor and the proofing time is unobservable, adjustable and variable. The other factors are considered as given constants. For simplicity, we assume that the random part of proofing time is normally distributed with variance 1. Experience about the natural variability should be used to determine this distribution in practice.

The data used for building the model relating protein content, mixing time and proofing time on one side to end product quality on the other, was based on a designed experiment. Three wheat flour qualities were mixed according to a $\{3,3\}$ simplex-lattice design, giving ten different flour mixtures. For each flour mixture, the protein content was measured. The measured protein contents ranged from 10.2 % to 14.3 %. Each flour was baked at different process conditions; three levels of mixing time; 5, 15 and 25 minutes, and three levels of proofing time; 35, 47.5 and 60 minutes. The process conditions were set up in a full factorial design, giving nine different conditions for each flour. The levels of the process conditions covered the region of interest for the present product.

The combined design consisted of crossing each of the ten blends of the mixture design with each of the nine points of the factorial design, making up a total of 90 combinations. The 90 trials were performed completely at random. All blends in the experiment were taken from large batches of the three flour qualities. The flour batches were thoroughly homogenized before making the ten flour blends, and each flour blend was homogenized during mixing. This leads to a low variation due to lack of homogeneity and imprecise mixing of the flours, compared to the other experimental factors. This, together with the random performance of the 90 trials enabled the fitting of models containing only a single error term [21].

For every hearth bread produced, the volume and form ratio = height/width were recorded. Also, skilled bakers recorded their subjective evaluation of the form ratio of the loaves on a scale from 1 to 4, where 4 corresponds to the highest quality and 1 to the lowest. The data comes from a larger study, and is more fully described by Næs et al. [21], and Færgestad et al. [22].

Second degree polynomials combined with variable selection to obtain a parsimonious model were used for modeling. Details will be given below.

5.1 Example 1: Single-Response Optimization

In the first example, we assumed that loaf volume is the only quality characteristic of interest, and that it has a finite target value, $v_t = 530$ ml. This target value was obtained from bakers' subjective judgments of the loaves. For larger volumes other, less attractive properties became apparent [21]. We found the volume to be related to the process conditions and the protein content through the model

$$\hat{v} = 523.3 + 2.0x_1 + 4.5x_2 + 22.9p - 0.30x_1^2 - 0.077x_2^2 + 0.84px_1 + 0.52px_2, \quad (6)$$

where x_1 represents mixing time and x_2 proofing time, centered around their sample means 15 and 47.5 minutes, respectively. The variable p represents protein content, and is centered around it's sample mean 12.4 %. We will use "p" and "protein content" interchangeably, although the protein content equals $p + 12.4$ %. Similarly for the other factors. Forward and backward variable selection, allowing up to second-degree terms, gave the same model. All coefficients were significant at a level of 0.05. The model had a squared multiple correlation coefficient R^2 of 0.88, and the root mean squared error of estimation (RMSEE) was 22.8 ml, with 82 degrees of freedom. Visual inspection of the residuals (not shown) showed no anomalies. The model was considered a reasonable approximation of the data.

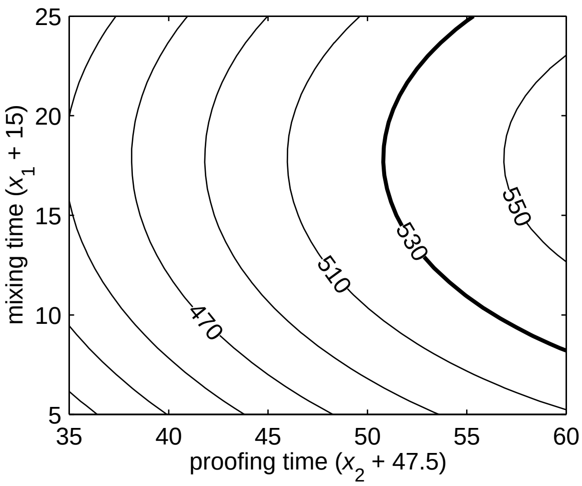

Fig. 1. Predicted loaf volume for 12 % protein. The thick contour line marks the target volume.

Note that for given process settings, the effect of protein is linear, and that the effects of both mixing time and proofing time are non-linear. Figure 1 shows a contour plot of the predicted volume for 12 % protein. The target volume is marked with a bold contour curve.

We chose a quadratic loss $(v - v_t)^2$ since no other information about the form of the actual loss was available. Using the predicted volume we get the loss function $L(x_1, x_2, p) = (\hat{v}(x_1, x_2, p) - v_t)^2$, a function of x_1, x_2 and p.

Note that $\hat{v}$ is an estimate of v, and L is therefore an *estimate* of the loss. Uncertainty due to random error and model error is thus propagated to L. This could be taken into account when comparing the predicted loss at different settings of the adjustable factors. One could identify a *region* around the calculated optima, or give a confidence interval of the actual loss at these points. This is not pursued further here, but will be treated in a forthcoming paper.

We calculated the expected loss $\mathsf{E}_{x_2}[L(x_1, x_2, p)]$ by taking the expectation over the variation in proofing time. The expected loss is then a function of protein content, p, mixing time, x_1, and the nominal setting of proofing time, μ_2.

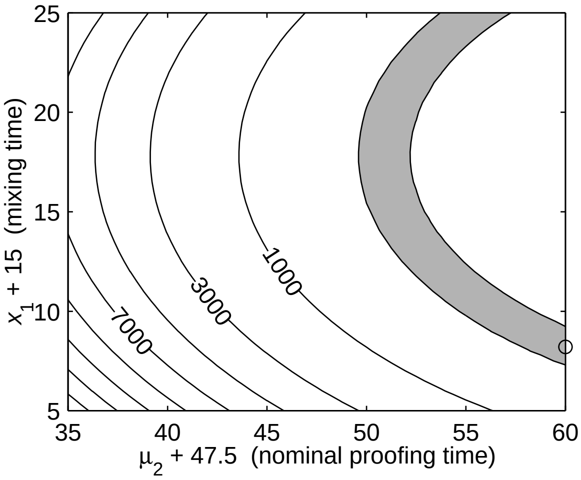

Fig. 2. Expected quadratic loss for 12 % protein. The circle marks the minimal point and the shaded region represents the lower 0.2 % of the range of $\mathsf{E}\,L$.

Figure 2 plots the expected loss for 12 % protein as a function of x_1 and μ_2. The minimal point within the experimental region is marked with a circle. This is the optimal initial settings for flours with 12 % protein content, and gave an expected loss equal to 5.6. Taking the empirical nature and random noise of the modeling into account, it seems natural to look at regions as well as exact points. The shaded region is where the expected loss is less than 39.2. This is the lower 0.2 % of the range of the expected loss for 12 % protein. Although the minimal expected loss occurs in a single point, there is a region of process settings giving approximately the same expected loss.

Since the experimental ranges of mixing and proofing times are thought to include at least the whole region of interest for the present product, all optimizations were restricted to the experimental region, i.e., $x_1 \in [-10, 10]$ and $\mu_2 \in [-12.5, 12.5]$. This is also reasonable since we use estimated models, which may be invalid outside the experimental region.

Using numerical minimization we calculated the optimal nominal settings and the corresponding minimal expected loss, for p ranging from 10 % to 14.5 %. The optimal settings are shown in Figure 3. All the optimal settings lie on the edge of the allowed region, in the lower right corner of the region. Færgestad et al. [22] show that with less than 5 minutes mixing time, the dough will not develop properly, and that short mixing times combined with long proofing

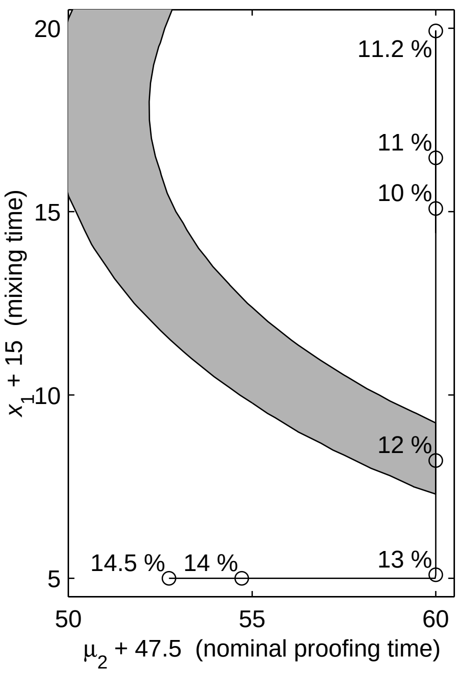

Fig. 3. Optimal mixing and proofing times for expected quadratic loss of volume, parameterized by protein. Part of the shaded region in Figure 2 is also shown.

times tend to produce loaves with too low form ratio. This indicates that the loaf volume alone might not be the best quality criterion.

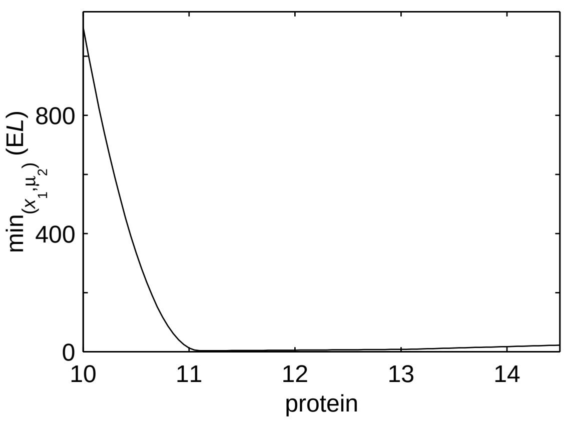

Fig. 4. Minimal expected quadratic loss of volume, as a function of p.

The minimal expected loss is plotted in Figure 4. There is a rapid increase in minimal expected loss as the protein content decreases below 11 %. The reason for this is that, according to the model (6), no process condition inside the experimental region gives volumes as high as 530 ml for flours with less than 11.1 % protein. This is consistent with the fact that no experimental run with lower protein content than 11.5 % produced volumes as high as 530 ml. The expected loss was close to zero for protein content between approximately 11 % and 11.5 %. This is because the maximal predicted volume is close to the target volume for these flours, and around the maximum the volume surface is quite flat. Thus small deviations in proofing time give little changes in volume, leading to a small expected loss.

5.1.1 Comparison with Ordinary Robust Optimization

In order to compare this batch-wise optimization with robust optimization that does not take the protein measurements into account, we treated the protein content as unobservable, and for simplicity assumed that it is uniformly distributed between 10.0 % and 14.5 %. We used the same loss function as before, but calculated the expected loss by taking the expectation over both proofing time and protein content. It is plotted in Figure 5. The optimal

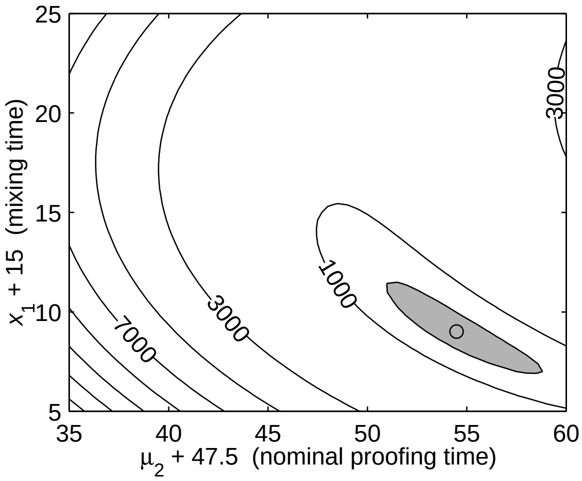

Fig. 5. Expected quadratic loss when p is considered unobservable. The circle marks the minimal point; where the expected loss was 823.1. The region where the expected loss is less than 0.2 % of its range, has been shaded.

settings of mixing and proofing time were 9.0 and 54.5 minutes, respectively, which gave a minimal expected loss of 823.1. This is the minimal expected loss as protein content and proofing time vary:

$$\mathsf{E}_{x_2,p}\left[L\left(x_1^{\mathrm{opt}}, x_2, p\right)\right], \quad x_2 \sim N\left(\mu_2^{\mathrm{opt}}, 1\right), \tag{7}$$

where x_1^{opt} and μ_2^{opt} are the optimal settings, functionally independent of p.

A relevant comparison of this optimization with the batch-wise robust optimization is to compare (7) with the expectation over p of the minimal expected loss for each p:

$$\mathsf{E}_p\left[\mathsf{E}_{x_2^p}\left[L\left(x_1^{\mathrm{opt}}(p), x_2^p, p\right)\right]\right], \qquad x_2^p \sim N\left(\mu_2^{\mathrm{opt}}(p), 1\right), \tag{8}$$

i.e., the expectation over p of the minimal expected losses in Figure 4. Here

$x_1^{\mathrm{opt}}(p)$ and $\mu_2^{\mathrm{opt}}(p)$ are the optimal settings for protein content p. The expectation was 98.5, which is a substantial reduction from 823.1. To further investigate the difference, we split the expected losses into sums of variance and squared bias. The resulting statistics are shown in Table 2.

	Unobservable protein content	Observable protein content
$\mathsf{E}\,L$ (RMSE)	823.1 (28.7 ml)	98.5 (9.9 ml)
Var (std)	793.2 (28.2 ml)	82.3 (9.1 ml)
Bias^2 ($\lvert\mathrm{Bias}\rvert$)	29.9 (5.5 ml)	16.2 (4.0 ml)

Table 2
Expected losses, variances and squared biases when protein is treated as unobservable or observable. The square roots are shown in parentheses. Results for 10 % $\leq$ protein $\leq$ 14.5 %.

We see that the reduction in expected loss was mainly due to reduction of variance, but the bias was reduced as well. The bias was not reduced further because no baking condition gave volumes as high as the target volume, for flours with less than 11 % protein.

Because of the rapid growth in minimal expected loss as the protein decreases below 11 % we performed the optimization using only flours with at least 11 % protein. This gave an average minimal expected loss of 9.9 for the batch-wise optimization and 447.3 when treating protein content as unobservable. In the last case the optimal settings were 8.3 minutes mixing time and 53.5 minutes nominal proofing time. The results of splitting the expected losses into sums of bias and variance are shown in Table 3. The response was practically unbiased, and the variance was reduced by a factor of approximately 45, when taking protein content into account.

	Unobservable protein content	Observable protein content
$\mathsf{E}\,L$ (RMSE)	447.2 (21.1 ml)	9.9 (3.1 ml)
Var (std)	439.9 (21.0 ml)	9.8 (3.1 ml)
Bias^2 ($\lvert\mathrm{Bias}\rvert$)	7.4 (2.7 ml)	0.00041 (0.020 ml)

Table 3
Expected losses, variances and (squared) biases when protein is treated as unobservable or observable. The square roots are shown in parentheses. Results for 11 % $\leq$ protein $\leq$ 14.5 %.

This kind of comparison is easy to do and should always be done in practice in order to assess the benefit of taking the raw material property into account when optimizing the process.

5.2 *Example 2: Multi-Response Optimization*

In the second example, we considered two quality characteristics: the loaf volume from the first example, and the form ratio of the loaves. We specified loss functions for each of them, and used a weighted sum of the loss functions as the criterion to be minimized.

For notational simplicity, we will in this example measure volume in liters, not milliliters. We standardized the expected losses before combining them, making them scale invariant, so this does not lead to a different solution.

We will first discuss the loss function used for each of the responses before we describe how the combined criterion is optimized.

5.2.1 *Volume*

The target value for volume was set to infinity in this example, so we decided to use the inverse quadratic loss function; $L_v = 1/\hat{v}^2$. It is commonly used when the target value is infinity, and is one of the losses suggested by Taguchi [3]. The expected loss with respect to the variation in proofing time, $\mathsf{E}_{x_2}[L_v(x_1, x_2, p)]$, is in this case not possible to represent in a closed form of simple functions, but it can be numerically computed for each combination of the independent variables.

For any p, the unrestricted settings of mixing and proofing time minimizing $\mathsf{E}\, L_v$ is the maximal point of the predicted volume $\hat{v}$. This can be shown to be a consequence of the shape of the loss function and the symmetrical distribution of x_2. The unrestricted maximal point of $\hat{v}$ can be found by solving $\partial\hat{v}/\partial x_1 = 0$ and $\partial\hat{v}/\partial x_2 = 0$ for x_1 and x_2. This gives the optimal settings

$$x_1^{\text{opt}} = 3.4 + 1.4p, \tag{9}$$

$$\mu_2^{\text{opt}} = 29.2 + 3.4p. \tag{10}$$

When the settings are confined to the experimental region, the optimal mixing time is equal to x_1^{opt} as above, while the optimal proofing time setting is equal to 60 minutes, or $\mu_2 = 12.5$, for all $p \in [10.0, 14.5]$.

Figure 6 shows a contour plot of the expected loss for 12 % protein. The region corresponding to the lower 0.2 % of the range is very small; cf. Figure 2. This means that the surface is steep near the optimum, compared to the expected loss surface in the first example, so a small change in nominal process settings can lead to a noticeable change in expected loss.

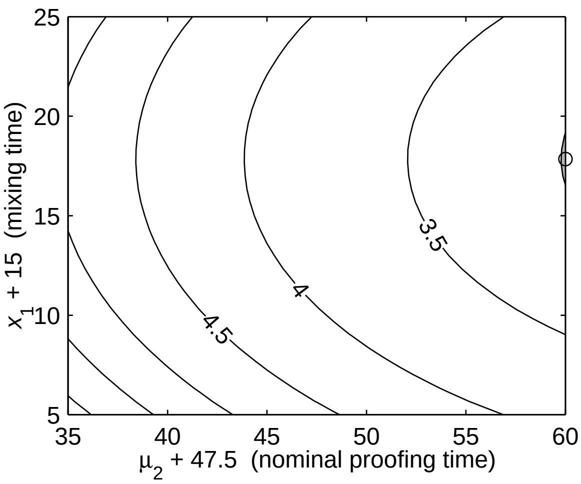

Fig. 6. Expected inverted quadratic loss of volume for 12 % protein. The minimal point, marked with a circle, lies on the axis $x_1 = x_1^{\text{opt}}$ of the paraboloid $\hat{v}$. The shaded region is where the loss is in the lower 0.2 % of its range; $\mathsf{E}\,L \leq 3.217$. It is barely visible around the minimal point.

5.2.2 Form Ratio

5.2.2.1 Defining a loss function when subjective assessments are available. We had access to subjective assessments of the form ratio, which could be used to define an appropriate loss function, by regression of the assessments on the measured form ratios. The logarithm of the assessments was used, since visual inspection suggested that the variance of the assessments was proportional to their mean. The data points are plotted in Figure 7. The

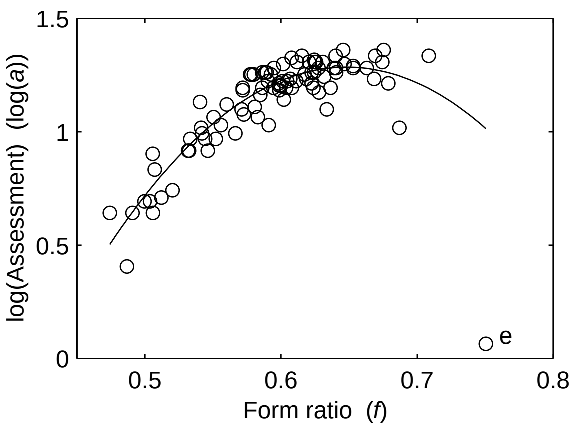

Fig. 7. Assessment of form ratio versus form ratio. Data points and regression curve. The point marked “e” was left out of the regression.

scatter plot shows substantial curvature, so a model of degree two or higher seemed appropriate. This is also supported by Færgestad et al. [22]. There was one extreme sample, marked “e” in the plot. This dough raptured during baking in oven due to large tensions when flour of high protein content was mixed for long time but proofed for a short time [22]. It had a large influence on the model estimation, so it was excluded from the regression. We found a significant second-degree model

$$\widehat{\log a} = 1.21 + 2.81 f_c - 25.8 f_c^2, \tag{11}$$

relating the assessment a to the form ratio f of the loaves. Here f_c is the form ratio centered around its sample mean $\bar{f} = 0.594$. Each term was significant at a level of 0.05. The model had an RMSEE of 0.076 with 86 d.f., and R^2 was 0.86. Both forward and backward variable selection allowing up to third degree terms gave this model.

$\widehat{\text{Log}\, a}$ can be written as a polynomial of the uncentered form ratio f:

$$y(f) = -9.57 + 33.5 f - 25.8 f^2. \tag{12}$$

This describes a parabola with a maximal value $y_{\max} = 1.29$ at $f_t = 0.648$. We specified the loss as

$$L_f = y_{\max} - y(f). \tag{13}$$

It is easy to show that $L_f = 25.8(f - f_t)^2$, so the loss is quadratic and the target value for the response is $f_t = 0.648$. Thus regression of the bakers' subjective assessments on the measured responses produced both a target value and a loss function. We believe that this represents a natural and interesting way of deriving loss functions, which can easily be extended to several responses.

5.2.2.2 Response Model and Expected Loss. The form ratio was found to be related to protein and the process condition through a second degree model

$$\begin{aligned}\hat{f} = {} & 0.61 + 3.8 \times 10^{-3} x_1 - 3.0 \times 10^{-3} x_2 + 8.3 \times 10^{-3} p \\ & - 2.9 \times 10^{-4} x_1^2 + 6.5 \times 10^{-5} x_2^2 + 7.0 \times 10^{-5} x_1 x_2 + 7.9 \times 10^{-4} p x_1. \end{aligned} \tag{14}$$

The variables are centered around their sample means. Each term was significant at a level of 0.05. For this model, $R^2 = 0.86$, and RMSEE $= 0.020$ with 82 degrees of freedom. Both forward and backward variable selection, allowing up to the full second-degree model gave the same model. Inspection of the residuals showed that the sample marked "e" in Figure 7 had the largest residual. However, since the sample had only a modest influence on the estimated model for the form ratio, it was kept in the current regression, even though it was removed from the regression for the assessments. Figure 8 shows a contour plot of the form ratio for 12 % protein.

Note that the second degree and product terms of the model only contain terms already present in the linear part of the model. Like the previous models (6) and (11), it therefore satisfies the requirements of marginality given in [23].

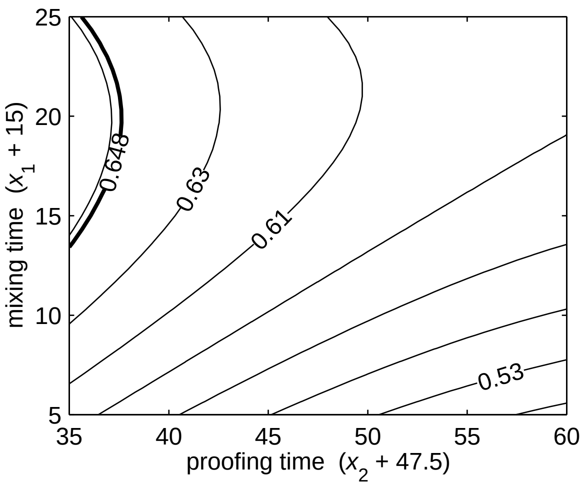

Fig. 8. Predicted form ratio for 12 % protein. The thick contour curve marks the target value.

The expected loss $\mathsf{E}_{x_2}[L_f(x_1, x_2, p)]$ is a fourth degree polynomial in x_1 and μ_2, with up to second-degree polynomials in p as coefficients. It is plotted for 12 % protein in Figure 9.

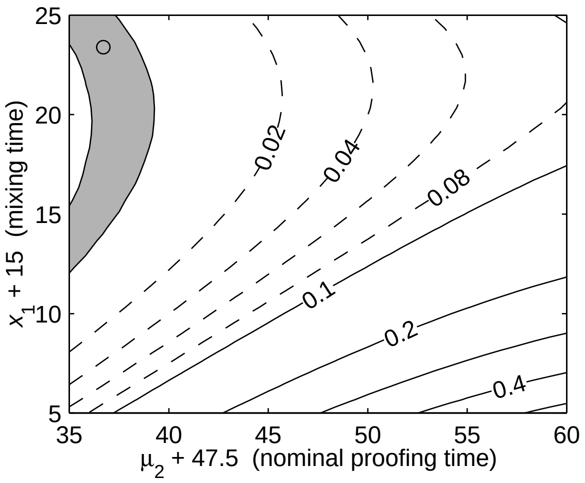

Fig. 9. Expected quadratic loss of form ratio, for 12 % protein. The circle marks the minimal point. The shaded region is the area where the loss is in the lower 0.2 % of its range; $\mathsf{E}\,L \leq 1.46 \times 10^{-3}$.

The optimal settings for the form ratio are located in the upper left corner of the region of interest, and for most p they lie on the border of the region. This suggests that form ratio alone might not be a suitable quality criterion.

5.2.3 Combining the Loss Functions

We used a weighted sum of the two loss functions to form a total loss, with weights 10 for form ratio and 1 for volume. The weights were chosen to reflect that form ratio was considered more important than volume. In order to make the loss functions scale invariant, they were standardized by dividing them by their standard deviation, calculated over the (x_1, x_2, p) parameter space. When doing this, we treated the mixing time, proofing time and protein content as independent variables, uniformly distributed on $[5, 25]$, $[35, 60]$ and $[10.0, 14.5]$, respectively. Other distributions could be used, and information regarding the distribution of the parameters during production could be incorporated. The total expected loss was

$$\mathsf{E}_{x_2}[L_{\text{tot}}] = 10 \times \mathsf{E}_{x_2}[L_f/0.099] + 1 \times \mathsf{E}_{x_2}[L_v/0.74]. \tag{15}$$

A contour plot for 12 % protein is shown in Figure 10. The minimal expected loss was 6.16.

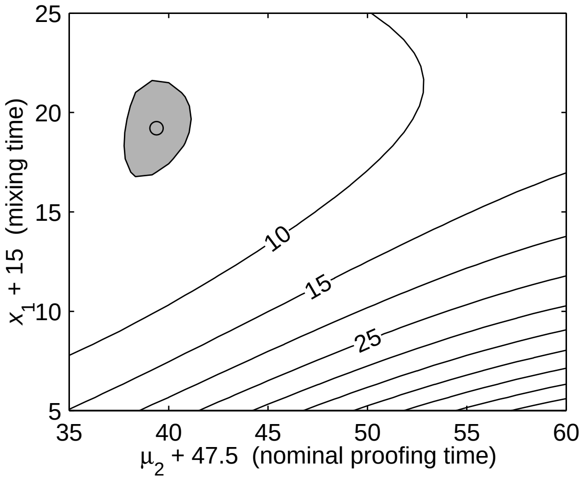

Fig. 10. Expected total loss for 12 % protein. The circle marks the optimal point. The region representing the lower 0.2 % of the range of $\mathsf{E}\,L$ is shaded; $\mathsf{E}\,L \leq 6.27$.

Constrained minimization of the expected total loss, restricting the mixing and nominal proofing times to the region of interest, gave optimal process settings as shown in Figure 11.

As in the first example, we then compared the batch-wise optimization with ordinary robust process optimization. The expected loss of the ordinary optimization was 6.51. The corresponding expectation for the batch-wise optimization, (8), was 5.95. The batch-wise optimization gave slightly smaller expected total loss for protein contents around 12 %, and substantial reductions for flours with low or high protein content.

5.2.4 Constrained Optimization

We also performed another optimization, this time under the constraint that the expected form ratio should be equal to its target 0.648. No process con-

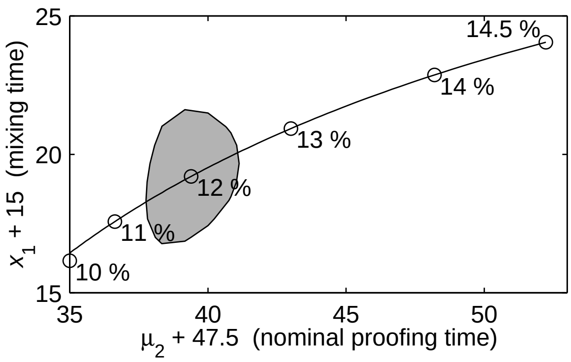

Fig. 11. Optimal settings for expected total loss, parameterized by protein. The shaded area is the region in Figure 10 corresponding to the lowest $\mathsf{E}\,L$ for 12 % protein.

dition gave high enough expected form ratio for flours with less than 11.05 % protein, so only flours with at least 11.05 % protein were used in the optimization. Other approaches are possible, for instance to modify the optimization criterion when the constraint is unattainable.

We compared batch-wise robust process optimization with ordinary robust process optimization under this constraint. The minimal expected losses were almost identical; 6.2 when using ordinary optimization versus 6.1 for batch-wise optimization. Batch-wise optimization gave lower variance of form ratio, but higher variance of volume. The mean volumes and form ratios were practically unchanged.

The reason why the results were so similar appears to be that the constraint on the form ratio leaves little room for batch-wise robust process optimization to improve on the results of ordinary robust process optimization. However, the constraint in practice makes the form ratio a primary response, and the volume secondary. The batch-wise optimization was able to reduce the variance of the primary response form ratio, sacrificing some of the robustness of the secondary response volume.

The main difference between the results of this constrained optimization and the unconstrained one above is that the unconstrained optimization allows the mean form ratio to be slightly off-target in order to increase the volume of the loaves, while the constrained optimization does not allow this. This results in a higher expected loss for the constrained optimization than for the unconstrained. A compromise between the two approaches could be achieved by allowing the mean form ratio to vary within an interval around the target value.

6 Conclusions

In this paper we have developed a unified description of the factors of a production process. All factors are characterized as stochastic variables, whose distributions we have total, some, or no control over. They are also classified as observable or unobservable depending on whether their measured values can be used in the robust optimization or control procedure. This description provides a good framework for understanding the process variation and for applying or developing robust optimization or control techniques.

We have proposed a robust process optimization technique for batch-wise processes. This batch-wise robust process optimization uses all available raw material measurements at the start of each production batch when calculating the optimal process settings for the batch. We have seen that this can lead to substantial reduction of loss due to unwanted variation, compared to ordinary robust process optimization, where the process is optimized once and for all.

In this study we have chosen a modeling approach and used loss functions. However, the general idea of using a batch-wise optimization and incorporating all available measurements for each batch, can equally well be applied to other approaches, such as the Taguchi techniques.

For multi-response systems, we have seen that the choice of optimization criterion is essential, and different criteria can lead to different solutions. This is an area where knowledge of the production process is very important, and an area for future research.

References

[1] George E. P. Box and Alberto Luceño. *Statistical Control by Monitoring and Feedback Adjustment.* Wiley Series in Probability and Statistics. Wiley, New York, 1997.

[2] Vijayan N. Nair. Taguchi's parameter design: A panel discussion. *Technometrics*, 34(2):127–161, 1992.

[3] Raghu N. Kackar. Off-line quality control, parameter design, and the taguchi method. *Journal of Quality Technology*, 17(4):176–209, 1985.

[4] Ramon V. León, Anne C. Shoemaker, and Raghu N. Kackar. Performance measure independent of adjustment: An explanation and extention of taguchi's signal-to-noise ratios. *Technometrics*, 29(3):253–285, 1987.

[5] Ramon V. León and C. F. Jeff Wu. A theory of performance measures in parameter design. *Statistica Sinica*, 2:335–358, 1992.

[6] G. Taguchi. *Experimental Design*, volume 1. Maruzen, Tokyo, 1976. In Japanese.

[7] G. Taguchi. *Experimental Design*, volume 2. Maruzen, Tokyo, 1977. In Japanese.

[8] Ravindra Khattre. Robust parameter design: A response surface approach. *Journal of Quality Technology*, 28(2):187–198, 1996.

[9] Raymond H. Myers, Andre I. Khuri, and G. Geoffrey Vining. Response surface alternatives to the taguchi robust parameter design approach. *The American Statistician*, 46(2):131–139, 1992.

[10] George E. P. Box and Stephen Jones. Designing products that are robust to the environment. *Total Quality Management*, 3(3):265–282, 1992.

[11] Enrique Del Castillo. Multiresponse process optimization via constrained confidence regions. *Journal of Quality Technology*, 28(1):61–70, 1996.

[12] Yves L. Grize. A review of robust process design approaches. *Journal of Chemometrics*, 9(4):239–262, 1995.

[13] J. Stuart Hunter. Statistical design applied to product design. *Journal of Quality Technology*, 17(4):210–221, 1985.

[14] George E. P. Box, William G. Hunter, and J. Stuart Hunter. *Statistics for Experimenters: an Introduction to Design, Data Analysis, and Model Building.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1978.

[15] John A. Cornell. *Experiments with Mixtures: Designs, Models and the Analysis of Mixture Data.* Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 2nd edition, 1990.

[16] George E. P. Box, Søren Bisgaard, and Conrad A. Fung. An explanation and critique of taguchi's contributions to quality engineering. *Quality and Reliability Engineering International*, 4:123–131, 1988.

[17] G. Taguchi. Off-line and on-line quality control systems. *Proceedings of International Conference on Quality Control*, 1978.

[18] Jan Engel and A. Freek Huele. Taguchi parameter design by second-order response surfaces. *Quality and Reliability Engineering International*, 12(2):95–100, 1996.

[19] Megan Pledger. Observable uncontrollable factors in parameter design. *Journal of Quality Technology*, 28(2):153–162, 1996.

[20] B. G. Osborne and T. Fearn. *Near Infrared Spectroscopy in Food Analysis.* Longman Science and Technical, Great Britain, 1986.

[21] Tormod Næs, Ellen Mosleth Færgestad, and John A. Cornell. A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables. *Chemometrics and Intelligent Laboratory Systems*, 41(2):221–235, 1998.

[22] Ellen Mosleth Færgestad, Ellen Merethe Magnus, Stefan Sahlström, and Tormod Næs. Influence of flour quality and baking process on hearth bread characteristcs made using gentle mixing. *Journal of Cereal Science*, 30:61–70, 1999.

[23] J. A. Nelder. Functional marginality and response-surface fitting. *Journal of Applied Statistics*, 27(1):109–112, 2000.