PD ISO/TR 13195:2015

BSI Standards Publication

Selected illustrations of response surface method — Central composite design

bsi.

...making excellence a habit.™

National foreword

This Published Document is the UK implementation of ISO/TR 13195:2015.

The UK participation in its preparation was entrusted to Technical Committee MS/6, Methodologies for business process improvement using statistical methods.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

ISBN 978 0 580 74473 0

ICS 03.120.30

Compliance with a British Standard cannot confer immunity from legal obligations.

This Published Document was published under the authority of the Standards Policy and Strategy Committee on 31 December 2015.

Amendments issued since publication

Date	Text affected

PD ISO/TR 13195:2015

TECHNICAL REPORT

ISO/TR 13195

First edition
2015-12-15

Selected illustrations of response surface method — Central composite design

Illustrations choisies de méthodologie à surface de réponse — Plans composites centrés

Reference number
ISO/TR 13195:2015(E)

COPYRIGHT PROTECTED DOCUMENT

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 69, *Applications of statistical methods*, Subcommittee SC 7, *Applications of statistical and related techniques for the implementation of Six Sigma*.

Introduction

The present Technical Report takes one specific statistical tool (Central Composite Designs in Response Surface Methodology) and develops the topic somewhat generically (in the spirit of International Standards) but then illustrates it through the use of four detailed and distinct applications. The generic description focuses on the Central Composite Designs.

The annexes containing the four illustrations follow the basic framework but also identify the nuances and peculiarities in the specific applications. Each example offers at least one "wrinkle" to the problem, which is generally the case for real applications. It is hoped that practitioners can identify with at least one of the four examples, if only to remind them of the basic material on response surface method that was encountered during their training.

Each of the four examples is developed and analysed using statistical software of current vintage. The explanations throughout are devoid of mathematical detail—such material can be readily obtained from the many design and analysis of experiments textbooks (such as those given in References [1] to [7]).

TECHNICAL REPORT **ISO/TR 13195:2015(E)**

Selected illustrations of response surface method — Central composite design

1 Scope

This Technical Report describes the steps necessary to understand the scope of Response Surface Methodology (RSM) and the method to analyse data collected using Central Composite Designs (CCD) through illustration with four distinct applications of this methodology.

Response surface methodology (RSM) is used in order to investigate a relation between the response and the set of quantitative predictor variables or factors. Especially after specifying the vital few controllable factors, RSM is used in order to find the factor setting which optimizes the response.

2 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

2.1
experiment
purposive investigation of a system through selective adjustment of controllable conditions and allocation of resources

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.1. (The notes are not reproduced here.)

2.2
response variable
variable representing the outcome of an *experiment* (2.1)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.3. (Except for NOTE 3 the notes are not reproduced here.)

Note 2 to entry: A common synonym is “output variable”.

Note 3 to entry: The response variable is likely to be influenced by one or more *predictor variables* (2.3), the nature of which can be useful in controlling or optimizing the response variable.

2.3
predictor variable
variable that can contribute to the explanation of the outcome of an *experiment* (2.1)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.4. (The notes are not reproduced here.)

Note 2 to entry: Natural predictor variables are expressed in natural units of measurement such as degrees Celcius (°C) or grams per liter, for example. In RSM work, it is convenient to transform the natural variables to coded variables which are dimensionless variables, symmetric around zero and all with the same spread.

2.4
model
<experiment> formalized representation of outcomes of an *experiment* (2.1)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.2. (The notes and examples are not reproduced here except for NOTE 2 which is NOTE 1 in ISO 3534-3.)

Note 2 to entry: The model consists of three parts. The first part is the *response variable* (2.2) that is being modelled. The second part is the deterministic or the systematic part of the model that includes *predictor variable(s)* (2.3). Finally, the third part is the *residual error* (2.12) that can involve *pure random error* (2.13) and *misspecification error* (2.14). The model applies for the experiment as a whole and for separate outcomes denoted with subscripts. The model is a mathematical description that relates the response variable to predictor variables and includes associated assumptions. Outcomes refer to recorded or measured observations of the response variable.

Note 3 to entry: In some areas the term transfer function is used for the systematic part of the model.

EXAMPLE In the models considered in response surface methodology the deterministic or systematic part are polynomials in the predictor variables. A second order model with two predictor variables is written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon$$

where ε is the random error. The associated assumptions on the random error could be either that individual random errors are uncorrelated with constant variance or independent and normally distributed. The deterministic part of the model is the second degree polynomial in the predictor variables x_1 and x_2

$$\mathrm{E}y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

which explains the mean (Ey) of the response variable as a function of the predictor variables.

2.5
factor
<design of experiments> feature under examination as a potential cause of variation

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.5. (The notes are not reproduced here.)

Note 2 to entry: Generally the symbol k is used to indicate the number of factors in the experiment.

2.6
factor level
setting, value or assignment of a *factor* (2.5)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition (3.1.12). (The notes are not reproduced here.)

2.7
coding of factor levels
<design of experiments> one-to-one relabelling of factor levels

Note 1 to entry: The coding of factor levels facilitates the identification of the design and the properties of the design.

Note 2 to entry: In response surface experiments the actual (or natural or operational) levels are relabelled such that the coded levels are numeric and symmetric around 0.

Note 3 to entry: A two-level factor is usually coded to have coded levels −1 and +1. A factorial design where all factors are two-level factors can be coded such that all runs are represented as *factorial runs* (2.9).

Note 4 to entry: In central composite designs numeric (or continuous) factors with five levels are considered, except for the face-centred central composite deigns, where only three levels are needed, see note 6 to 2.7. If the actual (or natural or operational) levels are $l_1 < l_2 < l_3 < l_4 < l_5$ then the middle level l_3 shall be the average of the lowest level l_1 and the highest level l_5, and, furthermore, l_3 shall be the average of the intermediate levels l_2 and l_4. The form of the coding operation can be expressed as

$$\text{coded value} = \frac{\text{actual value} - l_3}{C}$$

where C is half the distance from l_2 to l_4. With this coding of the factors each *run* (2.8) of a central composite design can be identified as either a *factorial point* (2.9), a *centre point* (2.10), or an *star point* (2.11). This is the coding used in textbooks for discussing central composite designs.

Note 5 to entry: An alternative coding is sometimes applied in the computations in software programs. The form of the coding operation can be expressed as

$$\text{coded value} = \frac{\text{actual value} - l_3}{M}$$

where M is half the distance from the lowest level l_1 to the highest level l_5. This coding will be referred to as *software coding* in this Technical Report.

Note 6 to entry: In the face-centred CCD, only three levels of each factor are needed, so $l_1 = l_2 < l_3 < l_4 = l_5$, and l_3 shall be the average of the lowest level l_1 and the highest level l_5. This design could be of interest if it is difficult to select five levels of the factors. For the face-centred CCD, the possible coded values of a factor are only –1, 0, 1.The face-centred CCD is not rotatable, see 2.18.

Note 7 to entry: A class of designs that can be used to fit second order models and only require three equidistant levels of each factor are Box-Behnken designs. Box-Behnken designs are not central composite designs and are therefore not treated in this Technical Report. But they may be a useful alternative, if only three equidistant levels of each factor can be used, see References [5], [2] and [7].

2.8
run
experimental treatment
<design of experiments> specific settings of every *factor* (2.5) used on a particular *experimental unit* (2.15)

Note 1 to entry: Ultimately, the impact of the factors will be captured through their representation in the *predictor variables* (2.3) and the extent to which the model matches the outcome of the *experiment* (2.1).

EXAMPLE Consider a chemical process *experiment* (2.1) in which a high yield is the objective and the predictor variables are temperature, duration, and concentration of a catalyst. A run could be a setting of temperature of 350 °C, 30 min duration and 10 % concentration of the catalyst, assuming that all of these settings are possible and permissible.

Note 2 to entry: Adapted from ISO 3534-3:2013, definition 3.1.13.

2.9
factorial point
factorial run
cube point
cube run
vector of factor level settings of the form (a_1, a_2, ..., a_k), where each a_i equals –1 or +1 as a notation for the coded levels of the factors

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.37. (The notes are not reproduced here.)

2.10
centre point
centre run
vector of factor level settings of the form (a_1, a_2, ..., a_k), where all a_i equal 0, as notation for the coded levels of the factors

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.38. (The notes are not reproduced here.)

2.11
star point
axial point
star run
axial run
vector of *factor level* (2.7) settings of the form (a_1, a_2, …, a_k), where one a_i equals α or $-\alpha$ and the other a_i's equal 0, as notation for the coded levels of the *factors* (2.6)

Note 1 to entry: For a k factor experiment, this process yields $2k$-star points of the form: (±α, 0, …, 0), (0, ±α, 0, …, 0), …, (0, 0, …, ±α).

Note 2 to entry: Star points are added to the design in order to estimate a quadratic response surface.

Note 3 to entry: Special values of α give a nice geometric structure. For a k factor experiment, if $\alpha = \sqrt{k}$ then the factorial points and the star points are all on the sphere with radius $\sqrt{k}$. This design is therefore called a spherical CCD. If $\alpha = 1$, the star points are on the faces of the unit cube and the design is a face-centred CCD.

2.12
residual error
error term
random variable representing the difference between the *response variable* (2.3) and its prediction based on an assumed *model* (2.4)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.6. (The notes are not reproduced here.)

2.13
pure random error
pure error
part of the residual error (2.12) associated with replicated observations

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.9. (The notes are not reproduced here.)

2.14
misspecification error
part of the *residual error* (2.12) not accounted for by *pure random error* (2.13)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.9. (The notes are not reproduced here.)

2.15
experimental unit
<design of experiments> basic unit of the experimental material

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.24. (The notes are not reproduced here.)

2.16
designed experiment
experiment (2.1) with an explicit objective and structure of implementation

Note 1 to entry: The purpose of a properly designed experiment is to provide the most efficient and economical method of reaching valid and relevant conclusions from the experiment.

Note 2 to entry: Associated with a designed experiment is an *experimental design* (2.17) that includes the *response variable* (2.2) or variables and the *experimental treatments* (2.8) with prescribed *factor levels* (2.6). A class of models that relates the response variable to the predictor variables could also be envisaged.

Note 3 to entry: Adapted from ISO 3534-3:2013, definition 3.1.27.

2.17
experimental design
assignment of *experimental treatments* (2.7) to each *experimental unit* (2.15)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.28. (The notes are not reproduced here.)

2.18
rotatability
characteristic of a *designed experiment* (2.16) for which the *response variable* (2.2) that is predicted from a fitted *model* (2.4) has the same variance at all equal distances from the centre of the design

Note 1 to entry: A design is rotatable if the variance of the predicted response at any point ***x*** depends only on the distance of ***x*** from the *centre point* (2.10). A design with this property can be rotated around its centre point without changing the prediction variance at ***x***.

Note 2 to entry: Rotatability is a desirable property for *response surface designs* (2.25).

Note 3 to entry: Rotatability of a central composite design is obtained setting α equal to the fourth root of the number of factorial points, i.e

$$\alpha = (n_F)^{1/4}$$

where n_F denotes the number of factorial points in a CCD.

Note 4 to entry: The definition and notes 1 and 2 are adapted from ISO 3534-3:2013, definition 3.1.40.

2.19
interaction
influence of one *factor* (2.6) on one or more other factors' impact on the *response variable* (2.2)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.17. (The notes are not reproduced here.)

2.20
factorial experiment
designed experiment (2.16) with one or more *factors* (2.5) and with at least two levels applied for one of the factors

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.2.1. (The notes are not reproduced here.)

2.21
full factorial experiment
factorial experiment (2.12) consisting of all possible combinations of the levels of the *factors* (2.6)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.2.2. (The notes are not reproduced here.)

2.22
fractional factorial experiment
factorial experiment (2.12) consisting of a subset of the *full factorial experiment* (2.21)

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.2.3. (The notes are not reproduced here.)

2.23
randomization
process used to assign treatments to experimental units so that each experimental unit has an equal chance of being assigned a particular treatment

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.26. (The notes are not reproduced here.)

2.24
replication
performance of an experiment more than once for a given set of predictor variables

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.1.35. (The notes are not reproduced here.)

2.25
response surface design
designed experiment (2.16) that identifies a subset of *factors* (2.5) to be optimized

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.2.19. (The notes are not reproduced here.)

2.26
analysis of variance
ANOVA
technique which subdivides the total variation of a *response variable* (2.2) into components associated with defined sources of variation

Note 1 to entry: Adapted from ISO 3534-3:2013, definition 3.3.8. (The notes are not reproduced here.)

3 Symbols and abbreviated terms

3.1 Symbols

y	Response variable
$\hat{y}$	Predicted response variable
$\hat{y}_S$	Predicted response variable at the stationary point
$\boldsymbol{x}_S$	Stationary point of fitted response surface
D_S	Distance of stationary point to the design centre
A, B, C, D	Factors
k	Number of factors
2^k	Number of runs in a full factorial experiment with k factors all having two levels
2^{k-p}	Number of runs in a fractional factorial experiment with k factors and fraction 2^{-p}
n_F	Number of factorial points in a CCD
n_S	Number of star points in a CCD
n_0	Number of centre points in a CCD
a_i, b_i, l_i	Levels of factors
+1, –1	High and low coded factorial levels
$-\alpha$, α	Axial levels of coded factors
σ	Standard deviation

3.2 Abbreviated terms

ANOVA	analysis of variance
CCD	central composite design
DOE	design of experiments

RSM response surface methodology

R&R repeatability and reproducibility

4 Generic descriptions of central composite designs

4.1 Overview of the structure of the examples in Annexes A to D

This Technical Report provides general guidelines on the design, conduct and analysis of central composite designs consisting of a specified number of two-level factors, and illustrates the steps with four distinct applications given in the annexes. Each of the four examples in Annexes A through D follows the basic structure as given in Table 1.

Table 1 — Basic steps in CCD design

1	Overall objective(s) of experiment
2	Description of the response variable(s)
3	Identification of factors affecting the response(s)
4	Selection of levels for each factor
5	Identification of measurement systems
6	Layout plan of the CCD (depending upon which main effects and two factor interactions are to be studied) with "randomization" principle (if these are physical runs)
7	Analyse the results – numerical summaries and graphical displays
8	Present the results
9	Perform confirmation run

4.2 Overall objective(s) of a response surface experiment

Experiments may be conducted for a variety of reasons. Therefore, the primary objective(s) for the experiment should be clearly stated and agreed to by all parties involved in the design, conduct, analysis and implications of the experimental effort.

The main goal of response surface experiments is to create a model of the relationship between the factors and the response in order to explore optimum operating conditions. This involves choosing a design which allows the fitting of a quadratic function as the systematic part of the model. The Central Composite Design (CCD) can achieve this and this design has been popular since its introduction in the first paper on response surface methods in 1951.[1]

Although the fundamental method for fitting first order (linear) or second order (quadratic) function of the predictor variables to the response is regression, the focus is not on the individual regression coefficients but on the regression function, the response surface, as a whole. This emphasis is reflected in the name Response Surface Methodology. Strong arguments in favour of this approach are given on pages 508-509 of Reference [2].

Typically, the primary goal for the experiment is to find optimal operating conditions based on the estimated response surface, this could involve doing several experiments, using the results of one experiment to provide direction for what to do next. This next action could be to focus the experiment around a different set of conditions, or to collect more data in the current experimental region in order to fit a higher-order model or confirm what seemed to be the conclusion.

The CCD is an appropriate name because three types of design points can be identified after a coding of the factor levels: centre points (2.10), factorial points (2.9) and star points (2.11), and those design points are indeed centred at the origin of the design space after the coding of the factor levels (2.7).

Response surface experiments traditionally involve a small number of continuous factors. Some software packages have an upper limit of 8 factors. Response surface experiments are typically used when the investigators already know which factors are important. One way to obtain this knowledge is to apply a screening experiment, for example a fractional factorial experiment as explained in ISO/TR 12845.[11]

4.3 Description of the response variable(s)

Associated with the objective of an experiment is a continuous outcome or performance measure. A response of interest could involve maximization (larger is better), minimization (smaller is better) or meet a target value (be close to a specified value), but, in all cases, that task is one of optimization.

The response variable (denoted by the variable y) should be closely related to the objective of the experiment. For some situations, there are more than one variable of interest to be considered, although, typically, only a primary response variable will be associated with the experiment. In other cases, multiple responses should be considered. In case of multiple responses, the approach taken in response surface methodology is to analyse and optimize each response separately. The fitted response surfaces will then be studied to find settings that meet the requirements of all the responses. The example in Annex C has three responses.

4.4 Identification of measurement systems

Assessment of repeatability and reproducibility of the measurement systems for factors and responses should be done prior to designing the experiment.

4.5 Identification of factors affecting the response(s)

Response surface experiments are usually not done in isolation. They rely on prior knowledge concerning important influential variables on the selected response. If this knowledge is not available, it is necessary to conduct a different type of experiment to identify the factors affecting the response.

During the final selection of factors, attention shall be paid to the ability to set the levels of each individual factor independently of the other factors.

4.6 Selection of levels for each factor

There are two aspects to the selection of factors. One is selecting the *experimental region* which is the multidimensional range of interest for the factors selected. The other is the exact selection of the factor levels in such a way that the design has desirable properties. The first one requires subject matter knowledge as to the impact of factors on the response. The second one is more straightforward once the factors and the type of design to be used are known. The second one is further discussed in this Clause.

The response surface methods considered in this Technical Report are about the second order centre models using the CCD. The CCD is an augmentation of 2^k factorial experiments (or 2^{k-p} fractional factorial experiments). In addition to the two factorial levels that are used in the (fractional) factorial experiments, the user selects three additional levels, one centre level which is the average of the two factorial levels, and two extreme levels which are chosen symmetrically around the centre level and typically outside the range of the two factorial levels.

When the experimenter selects the levels of each factor he will be thinking in terms of the operational levels of a factor, the exact setting of a temperature, for example. But when studying the properties of the design and also when analysing the data from the design coded levels of the design are used.

If the actual (or operational) levels are $l_1 < l_2 < l_3 < l_4 < l_5$ then the middle level l_3 shall be the average of the lowest level l_1 and the highest level l_5, and, furthermore, l_3 shall be the average of the intermediate levels l_2 and l_4. The form of the coding operation can be expressed as Formula (1):

$$\text{coded value} = \frac{\text{actual value} - l_3}{C} \tag{1}$$

where C is half the distance from l_2 to l_4. The coded value of the upper extreme level, l_5, will be denoted by α, and the coded value of the lower extreme level, l_1, will be denoted by $-\alpha$. It is very important to note that the value of α is the same for all the factors of the design. Thus, the coded levels of all the factors are $(-\alpha, -1, 0, 1, +\alpha)$.

An alternative coding is sometimes applied in the computations in software programs. The form of the coding operation can be expressed as Formula (2):

$$\text{coded value} = \frac{\text{original value} - l_3}{M} \tag{2}$$

where M is half the distance from the lowest level l_1 to the highest level l_5.or, equivalently, the distance from l_3 to l_5. This coding will be referred to as *software coding* in this Technical Report.

When the levels of the factors have been chosen, the levels of the individual factors have to be combined to define the runs of the experiment. A CCD has three types of experimental runs: factorial, centre and axial ones.

4.6.1 Factorial runs

A factorial run is a setting of all k factors to coded levels either –1 or +1. The factorial runs are the runs used in 2^{k-p} fractional factorial experiments or 2^k factorial experiments.

Written as a k-dimensional vector in coded levels, the factorial run has the form (±1, ±1,..., ±1,..., ±1). Considered as points in k-dimensional space, the factorial runs are the vertices of a cube and the factorial runs are for this reason also called cube points.

There are 2^k different factorial runs with k factors.

4.6.2 Star runs

The star runs are those where one of the factors has its coded levels either $-\alpha$ or α and the remaining factors are at their coded level 0.

Written as a k-dimensional vector in coded levels, the star run has the form (0, 0,..., ±α,..., 0), having $-\alpha$ or α on the i^{th} position and 0 on all other positions.

Viewed as points in k-dimensional space, the star runs are located on the coordinate axes, and for this reason, the star runs are also called axial runs.

There are 2^k different star runs with k factors.

4.6.3 Centre run

The centre run is the one where all the factors are on their coded level 0. Written as a k-dimensional vector, it is the point (0, 0,..., 0).

There is only one centre run but the centre run may be replicated in a CCD. One reason for replicating the centre point is to get an estimate of pure error which can be used to check the fit of the model.

4.7 Layout plan of the CCD with randomization principle

In a report, the full description of the design and the observed responses should be given. In addition to reporting the levels of the factors, as described in 4.6, this includes reporting the following:

— the number of replications of the three types of design points;

— the number of blocks;

— the randomization.

If the design has only a few factors, a table where each run is represented as a row is useful for this purpose. It is easier to grasp the design if coded levels are used. The randomization can be explained by including a column giving the order in which the runs have been performed. If blocking is applied, a column can similarly be added to explain the allocation of runs to blocks.

As an illustration, Table 3 provides a basic layout of a CCD for 2 factors with full factorial design in *serial order*. While there is a standard order for factorial designs there is no such thing for a CCD, so the term *serial order* is used for a way to write the runs in a CCD in such a way that the design is easily recognized. The serial order used in Table 3 and throughout this Technical Report lists the factorial runs in standard order, first, followed by the star runs and, finally, the centre runs. Each row of the table represents one set of experimental conditions that when run will produce a value of the response variable *y*. The two factors are designated as A and B. In this case, only one response variable is shown, but more columns should be added if more than one response variable is studied. In this case, it is readily seen that the design has five centre runs replicated. The last column shows the *run order*, i.e. the order in which the experimental conditions have been applied. This experiment has been performed in one block.

Table 3 — Layout of a generic Central Composite Design

Serial order	A	B	y	Run order
1	–1	–1	y_1	6
2	1	–1	y_2	4
3	–1	1	y_3	13
4	1	1	y_4	7
5	–1,41	0	y_5	12
6	1,41	0	y_6	8
7	0	–1,41	y_7	9
8	0	1,41	y_8	5
9	0	0	y_9	2
10	0	0	y_{10}	10
11	0	0	y_{11}	1
12	0	0	y_{12}	11
13	0	0	y_{13}	3

4.8 Analyse the results — Numerical summaries and graphical displays

The second order model that can be fitted with the data from a CCD is a regression model and the usual output and graphs from a regression analysis are reported. This includes the table of estimated regression coefficients and associated t-statistics and *p*-values. A special feature of RSM regression is ANOVA tables that exploit the structure of the model.

For the design in Table 3, one such ANOVA table is shown in Table 4.

Table 4 — ANOVA table for Central Composite Design

Response: y					
	Df	Sum Sq	Mean Sq	F-value	Pr(>F)
FO(x1, x2)	2	ddd,dd	ddd,dd	ddd,dd	d,dd
TWI(x1, x2)	1	ddd,dd	ddd,dd	ddd,dd	d,dd
PQ(x1, x2)	2	ddd,dd	ddd,dd	ddd,dd	d,dd
Residuals	7	ddd,dd	ddd,dd		
Lack of fit	3	ddd,dd	ddd,dd	ddd,dd	d,dd
Pure error	4	ddd,dd	ddd,dd		

Table 4 is given by most, if not all, statistical packages for RSM analysis. The names of the rows and the columns may differ between statistical packages, but they are easily understood. Here, FO, TWI and PQ are short for First Order, Two Way Interaction and Pure Quadratic. Examples are given in Tables A.5, B.6 to B.7, and D.5.

In this case, there is no data, so the table is filled in with ddd,dd or d,dd except for the degrees of freedom which are the correct ones for the design in Table 3. The four F-tests are of interest. Read from the bottom of the table, the first one is the ratio of the mean squares for Lack of fit and Pure error and is a check of the fit of the model. The second F-test is in the PQ line and it tests the need for pure quadratic terms in the model; in this case the hypothesis that $\beta_{11} = \beta_{22} = 0$. It is the ratio of the mean square in the PQ line to the mean square in the Residuals line of the table.

Another ANOVA table of potential interest tests for each predictor variable, the hypotheses that it is not needed in the model. For the example with two predictor variables in Table 3, this ANOVA table has two F-tests: One for the hypothesis that $\beta_1 = \beta_{12} = \beta_{11} = 0$, and one for the hypothesis that $\beta_2 = \beta_{12} = \beta_{22} = 0$. Note that β_1, β_{12}, and β_{11}, for example, are the coefficients of all the terms in the model that involve x_1. Examples can be found in Tables B.5 and C.7.

When looking at analyses of response surface experiments using different software, it is possible to identify two different approaches. One approach uses all the tools of regression analysis to find a model where only significant terms are included in the estimated systematic part of the model. The other approach focuses on the response surface and only tests the hypotheses described in the two clauses above, i.e. tests whether a linear model can be used instead of a quadratic model or tests whether a factor can be considered to be essentially inert. In this Technical Report, the second approach is taken and therefore individual non-significant terms are not removed from the model. Strong arguments in favour of the second approach are given on pages 508-509 in Reference [2].

4.9 Present the results

Of course, one presentation of the result of the analysis is to give the estimated systematic part of the model. Although the systematic part of the model is indispensable for calculating the predicted response for various settings, it is not very useful for understanding the nature of the response surface. In Annex E, four second order polynomials in two variables are given in Formulae (E.3), (E.4), (E.5) and (E.6). The four polynomials look very similar. They have the same distribution of signs among the coefficients and the coefficients are very similar. But the perspective and contour plots in Figures E.1 to E.4 show that the polynomials represent four very different response surfaces.

For two and three predictor variables, it is possible to get an understanding of the response surface from contour plots, but a more formal analysis called *canonical analysis* is always useful, because it gives a precise characterization of the surface that gives an understanding of the response surface even when the number of predictor variables is larger than 3.

Canonical analysis is a method of rewriting a fitted quadratic function of the predictor variables in a form which can be more easily understood. This is achieved by a rotation of the coordinate axes which removes all cross-product terms. This may be followed by a change of origin to remove first order terms as well. Details are given in E.3.

Software packages for response surface methodology report the result of the canonical analysis. The result has the following parts:

— stationary point and predicted value at stationary point of the response surface;

— eigenvalues;

— eigenvectors.

The stationary point is the point where the fitted response surface has a maximum, a minimum or a saddle point, and the eigenvalues tell exactly which one is the case. Associated with each eigenvalue is an eigenvector. Moving away from the stationary point in the positive or negative direction of the eigenvector the response will decrease if the eigenvalue is negative, and it will increase if the eigenvalue is positive, so

— if all the eigenvalues are negative, the stationary point is a maximum,

— if all the eigenvalues are positive, the stationary point is a minimum, and

— if some eigenvalues are negative and some are positive, the stationary point is a saddle point or a minimax.

The situation is further complicated by the fact that one can only hope to approximate the response by the fitted second order model in the experimental region, so it should also be considered whether the stationary point is inside or outside the experimental region.

The situation is only simple, if the object is to maximize the response and if the stationary point is a maximum located inside the experimental region, or, alternatively, if the object is to minimize the response and if the stationary point is a minimum located inside the experimental region, and then the optimal settings is the stationary point. In all other cases, the optimal settings need to be found towards the boundary of the experimental region. The proprietary software packages provide a variety of tool for this purpose.

4.10 Perform confirmation run

After determining the optimal combination of levels of the factors and predicting the value of response variable for these factor levels, it is recommended that the user performs confirmation runs at the chosen settting to check whether the new observations confirm the predictions of the experiment.

If the computer output gives the standard deviation (std.dev) of a single observation and the standard error (std.err) of the predicted value at the optimum, then the following formulas apply for

a) a 95% confidence interval for the mean at the optimum: $\hat{y} \pm t_{0,975}(v) \cdot \text{std.err}$ and

b) a 95% prediction interval for a new observation at the optimum: $\hat{y} \pm t_{0,975}(v) \cdot \sqrt{\text{std.dev}^2 + \text{std.err}^2}$.

Here, v denotes the degrees of freedom for error.

The 95 % prediction interval is particularly useful for it explains what can be expected from future observations; if the model is good, 95 % of the observations will fall inside the 95 % prediction interval.

5 Description of Annexes A through D

5.1 Comparing and contrasting the examples

Four distinct examples of response surface designs are illustrated in Annexes A to D. Each of these examples follows the same general template as given in Table 1.

5.2 Experiment summaries

Table 5 summarizes the four examples detailed in the annexes and indicates aspects of the analyses which were unique to each experiment.

Table 5 — Experiment summaries

Annex	Experiment	Problem-specific aspects	Software used for analysis
A	Yield of a crop	Repetitions; important 2-factor interaction; contour plots	Minitab® 17[a]
B	Button tactility	Centre points; curvature in response	SAS® 9.4[a]
C	Semiconductor die deposition	Curvature in response; face-centred central composite design; three responses need to be optimized; trade-off between three responses	JMP® 10[a]
D	Palladium-copper catalysed C-C-bond formation	Replication; central composite design	destra®V11[a]

[a] This information is given for the convenience of users of this document and does not constitute and endorsement by ISO of these products. Equivalent products may be used if they can be shown to lead to the same results.

Annex A
(informative)

Effects of fertilizer ingredients on the yield of a crop[1)]

A.1 Purpose of the experiment

An experiment was conducted to investigate the effects of three fertilizer ingredients on the yield of beans under field conditions. The fertilizer ingredients applied were nitrogen (N), phosphoric acid (P_2O_5) and potash (K_2O). The response of interest was the yield in kg per plot.

Response surface methodology is often thought of as a sequential experimentation process which begins with a screening to determine the active factors. However, in this case, the active ingredients in the composition of the fertilizer were well understood and the purpose of the experiment was simply to find an optimal combination of the three ingredients. For this purpose, and because the surface was expected to be quadratic, a central composite design (CCD) was chosen.

A.2 Response variable

Only one response variable was considered in this experiment: Yield of beans measured in kg per plot.

A.3 Predictor variables

The three fertilizer ingredients were the predictor variables in the experiment. It was decided that the feasible amounts of the three fertilizer ingredients were within the ranges:

— **N**: Nitrogen: 0,425 kg/plot to 2,833 kg/plot

— **P_2O_5**: Phosphoric Acid: 0,266 kg/plot to 1,326 kg/plot

— **K_2O**: Potash: 0,278 kg/plot to 1,900 kg/plot

Special attention was given to the extreme settings and the scientist was confident that all combinations of extreme settings of the variables were within the region of operability.

All variables were continuous in the sense that all values within the chosen ranges were possible settings of the variables.

A.4 Identification and estimation of measurement systems

The repeatability and reproducibility of the measurement system for the associated response and predictor variables was deemed adequate for the objectives of this experiment, so no further work was initiated.

A.5 Selection of the settings of the predictor variables in the design

A.5.1 General considerations

To decide on the settings of the predictor variables in this experiment the following considerations were made. First, it was decided that the extreme levels of all variables should be the endpoints of the

1) Source: Renu Gupta, Bureau of Indian Standards, India.

ranges above. Taking nitrogen as an example and using the notation of 4.6 where the actual levels are denoted by $l_1 < l_2 < l_3 < l_4 < l_5$ implies that the lowest level is l_1 = 0,425, the highest level is l_5 = 2,833, and the middle level is $l_3 = (l_1 + l_5) / 2 = 1,629$. In order to find the values of l_2 and l_4, the value of α needs to be known. It was decided to use a rotatable design with k = 3 factors and 8 factorial points. By the formula in 2.18, Note 3, α is the fourth root of the number of factorial points, so $\alpha = 8^{1/4} = 1,682$. This means that α = 1,682 is the distance in coded units from the middle level to the highest level,

$$\frac{l_5 - l_3}{C} = 1,682 \text{ or } C = \frac{l_5 - l_3}{1,682} = 0,716 \tag{A.1}$$

Now, the actual cube levels l_2 and l_4 are found as

$$l_2 = l_3 - C = 1,629 - 0,716 = 0,913 \text{ and } l_4 = l_3 + C = 1,629 + 0,716 = 2,345 \tag{A.2}$$

Similar calculations were used on phosphoric acid and potash with the results given in Table A.1.

Table A.1 — Coded and actual levels of the experimental factors. Coded levels are in the top row and actual levels are in the following rows. The actual levels of all variables are kg/plot.

Coded levels	–1,682	–1,000	0,000	+1,000	+1,682
N	0,425	0,913	1,629	2,345	2,833
P_2O_5	0,266	0,481	0,796	1,111	1,326
K_2O	0,278	0,607	1,089	1,571	1,900

Coded values of N, P_2O_5 and K_2O are denoted by x_1, x_2 and x_3 respectively, and the conversions from the actual levels given in Table A.1 to the coded levels are given as

$$\begin{aligned} x_1 &= \frac{\mathrm{N} - 1,629}{0,716}, \\ x_2 &= \frac{\mathrm{P_2O_5} - 0,796}{0,315}, \\ x_3 &= \frac{\mathrm{K_2O} - 1,089}{0,482}. \end{aligned} \tag{A.3}$$

In a CCD, the settings or levels of the predictor variables are of three types to be considered next.

A.5.2 Factorial levels

The factorial levels are coded as –1 and +1, so the factorial levels in actual or natural units are given in the two columns labelled –1,000 and +1,000, in Table A.1. Factorial levels of the predictor variables in this experiment are the following:

— **N:** 0,913 and 2,345 (coded as –1 and 1 respectively);

— **P_2O_5:** 0,481 and 1,111 (coded as –1 and 1 respectively);

— **K_2O:** 0,607 and 1,571 (coded as –1 and 1 respectively).

The factorial levels have been given their name because those levels are used to define a factorial design, which is an important building block in a CCD. The factorial design is typically a full two-level factorial design or a fractional two-level design if the number of factors is large.

A.5.3 Axial levels

These are the extreme values for the predictor variables, so the axial levels in actual or natural units are given in the two columns labelled –1,682 and +1,682, in Table A.1. In this experiment, the axial levels are the following:

— **N:** 0,425 and 2,833 (coded as –1,682 and 1,682 respectively);

— **P_2O_5:** 0,266 and 1,326 (coded as –1,682 and 1,682 respectively);

— **K_2O:** 0,278 and 1,900 (coded as –1,682 and 1,682 respectively).

A.5.4 Centre levels

The centre levels are the average of the two factorial levels, or the average of the two axial levels. Its coded value is 0. In this experiment, the centre levels were the nominal values of the predictor variables. But this is not a requirement. The experimenter is free to choose the region of experimentation where the response surface is being investigated. In this experiment, the centre levels are the following:

— **N:** 1,629 (coded as 0);

— **P_2O_5:** 0,796 (coded as 0);

— **K_2O:** 1,089 (coded as 0).

A.6 Experimental design

A.6.1 General

The levels of the individual predictor variables defined in A.5 are combined to define the runs of the experiment. The three types of runs in the CCD are the factorial, the star (or axial) and the centre runs.

A.6.2 Factorial runs

In the factorial runs, all predictor variables are at their factorial levels. The factorial runs usually constitute a full factorial or a fractional factorial design. In this experiment, a full factorial design was chosen, so the number of factorial runs was $2^k = 2^3 = 8$. The factorial runs are given as serial numbers 1 to 8 in coded levels in Table A.2.

A.6.3 Axial runs

In the axial runs, one of the predictor variables is at its upper or lower extreme level and the rest of the predictor variables are at their centre value. The total number of axial runs in the experiment was $2k = 2 \times 3 = 6$ The axial runs are given as serial number 9 to 14 in coded levels in Table A.2.

A.6.4 Centre runs

In the centre runs, all of the predictor variables are at their centre level. The centre runs were replicated six times in this experiment. The six centre runs are given as numbers 15 to 20 in the serial order in Table A.2. This is an agricultural experiment where the experimental units are plots of land. The time order of the recording of the yields is not relevant in this case, and for this reason there is no run order given in Table A.2. It is important to randomize allocation of treatments to plots, but the information of the randomization has not been made available.

Table A.2 — Experiment layout in coded levels

Serial order	x_1	x_2	x_3
1	−1	−1	−1
2	1	−1	−1
3	−1	1	−1
4	1	1	−1
5	−1	−1	1
6	1	−1	1
7	−1	1	1
8	1	1	1
9	−1,682	0	0
10	1,682	0	0
11	0	−1,682	0
12	0	1,682	0
13	0	0	−1,682
14	0	0	1,682
15	0	0	0
16	0	0	0
17	0	0	0
18	0	0	0
19	0	0	0
20	0	0	0

A.7 Data generated by the experiment

The response variable from the experiment is given in the last column in Table A.3. The runs are given in coded levels in the first three columns and in actual levels in columns 4 to 6. The total amount of fertilizer applied is given in the seventh column and the yield is given in the last column.

Table A.3 — Data collected during the experiment. The Total Amount of Fertilizer is the row sum of N, P_2O_5 and K_2O columns.

x_1	x_2	x_3	N	P_2O_5	K_2O	Total amount of fertilizer	Yield
−1	−1	−1	0,913	0,481	0,607	2,001	5,076
1	−1	−1	2,345	0,481	0,607	3,433	3,798
−1	1	−1	0,913	1,111	0,607	2,631	3,798
1	1	−1	2,345	1,111	0,607	4,063	3,469
−1	−1	1	0,913	0,481	1,571	2,965	4,023
1	−1	1	2,345	0,481	1,571	4,397	4,905
−1	1	1	0,913	1,111	1,571	3,595	5,287
1	1	1	2,345	1,111	1,571	5,027	4,963
−1,682	0	0	0,425	0,796	1,089	2,310	3,541
1,682	0	0	2,833	0,796	1,089	4,718	3,541
0	−1,682	0	1,629	0,266	1,089	2,984	5,436

Table A.3 *(continued)*

x_1	x_2	x_3	N	P_2O_5	K_2O	Total amount of fertilizer	Yield
0	1,682	0	1,629	1,326	1,089	4,044	4,977
0	0	−1,682	1,629	0,796	0,278	2,703	3,591
0	0	1,682	1,629	0,796	1,900	4,325	4,693
0	0	0	1,629	0,796	1,089	3,514	4,563
0	0	0	1,629	0,796	1,089	3,514	4,599
0	0	0	1,629	0,796	1,089	3,514	4,599
0	0	0	1,629	0,796	1,089	3,514	4,275
0	0	0	1,629	0,796	1,089	3,514	5,188
0	0	0	1,629	0,796	1,089	3,514	4,959

A.8 Analysis of results

A.8.1 Software used for the analysis

Data acquired through the experiment are analysed using Minitab® 17.[2)]

A.8.2 Transfer function

The analysis was performed using the software coded versions of the predictor variables, but the estimated transfer function (the estimated systematic part of the model) for Yield of a crop (in kg) is more useful when expressed in the actual or natural units. The estimated transfer function in natural units in Table A.4 is taken from the Minitab® 17[2)] output.

Table A.4 — Estimated transfer function in actual units

Yield = 6,08 + 1,559 N - 6,01 P2O5 - 0,90 K2O - 0,739 N*N + 2,116 P2O5*P2O5 - 0,715 K2O*K2O - 0,142 N*P2O5 + 0,784 N*K2O + 2,411 P2O5*K2O

A.8.3 Estimation of coefficients

The result of the estimation is displayed in Table A.5 using software coded units.

As explained in the last paragraph of 4.8, the approach to analysis adopted in this Technical Report is the one recommended by Box et al. on pages 508-509 in Reference [2]. Thus, it is not recommended to inspect the individual regression coefficients and trying to interpret whether some of the coefficients might fail to be significantly different from 0. The point is that a second order surface is being fitted and the least squares estimate of the fitted surface has the coefficients in the "Estimated Regression Coefficients for Yield" part of Table A.4 or Table A.5 depending on the coding of the predictor variables used.

Table A.5 — Estimated regressions coefficients and CCD ANOVA table using software coded levels

Estimated regression coefficients for yield				
Term	**Coef**	**SE Coef**	**T**	**P**
Constant	4,692 60	0,145 4	32,282	0,000
N	−0,129 18	0,162 2	−0,796	0,444

2) This information is given for the convenience of users of this document and does not constitute and endorsement by ISO of these products. Equivalent products may be used if they can be shown to lead to the same results.

Table A.5 *(continued)*

P2O5	−0,130 21	0,162 2	−0,803	0,441
K2O	0,602 38	0,162 2	3,713	0,004
N*N	−1,070 92	0,265 5	−4,033	0,002
P2O5*P2O5	0,594 49	0,265 6	2,238	0,049
K2O*K2O	−0,470 01	0,265 6	−1,770	0,107
N*P2O5	−0,090 89	0,356 5	−0,255	0,804
N*K2O	0,765 69	0,356 5	2,148	0,057
P2O5*K2O	1,036 50	0,356 7	2,906	0,016
S = 0,356 408 PRESS = 7,324 51				
R-Sq = 84,40 % R-Sq(pred) = 10,07 % R-Sq(adj) = 70,37 %				

Analysis of variance for yield

Source	**DF**	**Seq SS**	**Adj SS**	**Adj MS**	**F**	**P**
Regression	9	6,874 26	6,874 26	0,763 81	6,01	0,005
Linear	3	1,913 51	1,913 51	0,637 84	5,02	0,022
Square	3	3,294 20	3,294 20	1,098 07	8,64	0,004
Interaction	3	1,666 54	1,666 54	0,555 51	4,37	0,033
Residual error	10	1,270 27	1,270 27	0,127 03		
Lack-of-Fit	5	0,745 30	0,745 30	0,149 06	1,42	0,355
Pure error	5	0,524 97	0,524 97	0,104 99		
Total	19	8,144 53				

While it is inappropriate to change the values of individual estimated regression coefficients, it does make sense to look for model simplification where one answers questions like the following:

— Are the pure quadratic terms needed in the model?

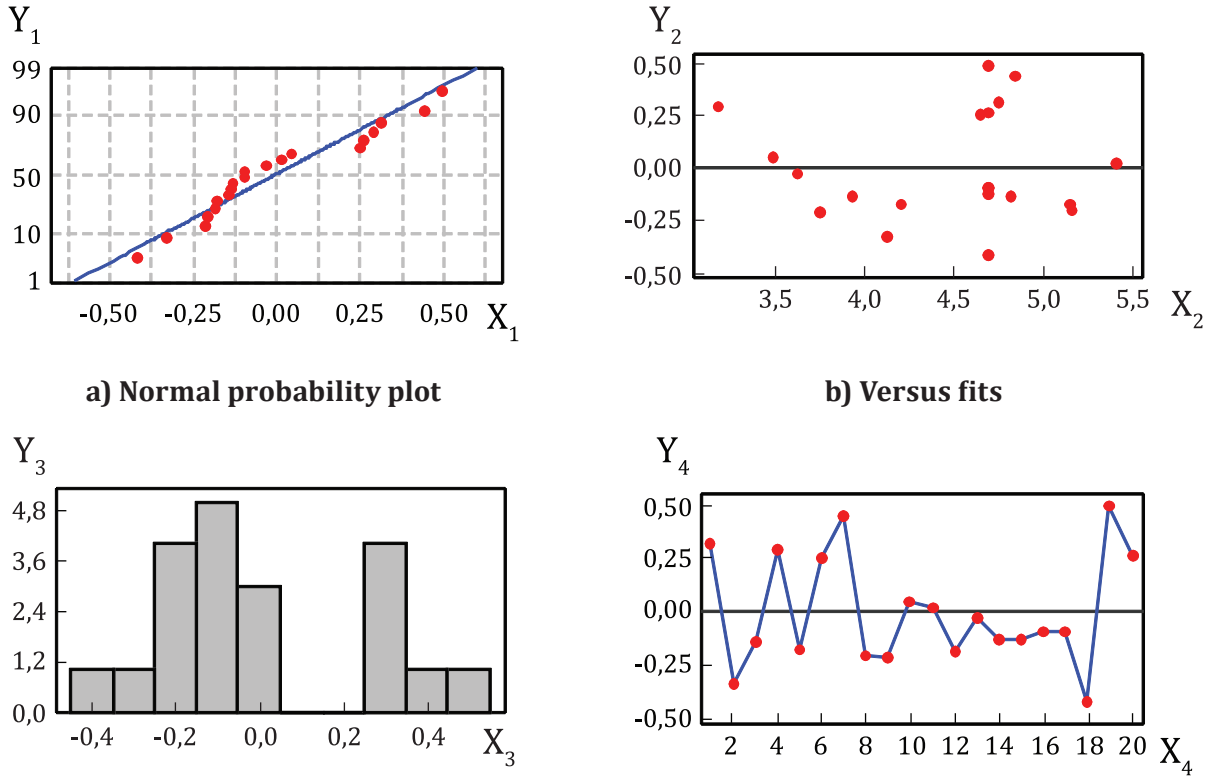
— Is a first order model satisfactory?

This is where the "Analysis of Variance for Yield" part of Table A.5 is useful. Read from the bottom of the table, the first F-test of interest is the ratio of the mean squares for Lack-of-Fit and Pure Error which comes from the replication of the centre points and is a check of the fit of the model. In this case, the *p*-value is 0,355 and the model is not questioned. The second F-test is in the Square line and it tests the need for pure quadratic terms in the model; in this case, the hypothesis that $\beta_{11} = \beta_{22} = \beta_{33} = 0$. It is the ratio of the mean square in the Square line to the mean square in the Residual Error line of the table. The *p*-value is 0,004 and the hypothesis is strongly rejected, so quadratic terms are needed in the model. The F-test in the Interaction line tests the hypothesis $\beta_{12} = \beta_{13} = \beta_{23} = 0$ and this hypothesis is also rejected and so is the hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$ which is tested by the F-test in the Linear line.

A.8.4 Graph of residuals

Figure A.1 shows some standard model checking plots based on the residuals that are produced by Minitab®. The observation order that is used in the plot of residuals against observation order is the serial order in Table A.2. There is a strange pattern in this plot: large residuals in the factorial runs, small residuals in the axial runs and the first three centre runs, and finally large residuals in the final three centre runs. A plot of residuals against observation order is most useful if observations are made in a natural order, for example a time sequence. As mentioned in A.6.4, the experiment is a field experiment so, no single one-dimensional plotting order is relevant. In this case the information about

how the plots were laid out in the field is not available, so the plot is not particularly relevant here but is kept as an illustration of a potentially useful plot.



a) Normal probability plot

b) Versus fits

c) Histogram

d) Versus order

Key

X_1	residual	Y_1	percent
X_2	fitted value	Y_2	residual
X_3	residual	Y_3	frequency
X_4	observation order	Y_4	residual

Figure A.1 — Residual plots

A.9 Presentation of results — Optimization

A.9.1 General

Just looking at the transfer function (the estimated systematic part of the model), whether it is shown in actual units as in A.8.2 or in coded units, does not reveal the properties of the response surface. It is important to realize that the second order polynomial in two or more variables can represent a variety of surfaces. The stationary point can be a maximum, a minimum or a saddle point. Furthermore, the stationary point can be located outside as well as inside the experimental region. The situation is only simple, if the object is to maximize the response and if the stationary point is a maximum located inside the experimental region, or, alternatively, if the object is to minimize the response and if the stationary point is a minimum located inside the experimental region, and then the optimal settings is the stationary point. In all other cases, the optimal settings need to be found towards the boundary of the experimental region. If the number of predictor variables is two or three, then, contour plots can be of help as illustrated in A.8.3. Alternatively, numerical methods, such as the desirability function, can be used. An application of the desirability function is given in A.9.2.

A.9.2 Application of the desirability function

The desirability function is most often presented as numerical tool that can solve conflicts where two or more responses have conflicting optimal settings, but it can also be used to find optimal settings with just one response as illustrated in this example.

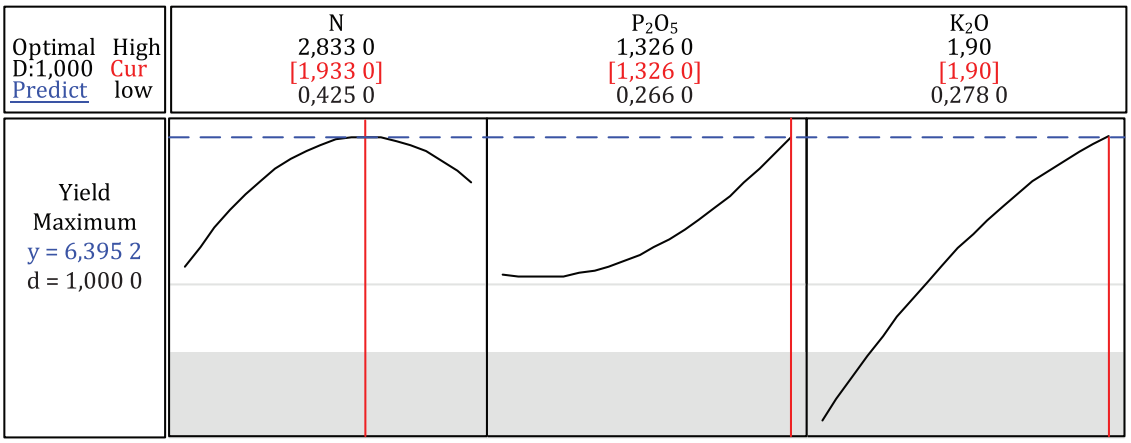


Figure A.2 — Optimization plot given by Minitab's® response optimizer

Using the response optimizer in Minitab® gives the optimization plot in Figure A.2. It indicates that a high value of yield can be achieved when N, P_2O_5 and K_2O are set at 1,933 0, 1,32 6 and 1,90 (values in Red), respectively. The estimated predicted value of yield at these values is 6,395 2 (value in blue).

Note that both P_2O_5 and K_2O are set at their extreme high values in the experiment. This means that the optimal setting suggested by Minitab's® response optimizer is rather far from settings that are used in the experiment. It could be argued that it lies outside the experimental region.

Clicking on Predict in the optimization plot in Minitab® gives useful information shown in Table A.6. It is very useful that both the 95 % confidence interval and the 95 % prediction interval for the predicted yield are given.

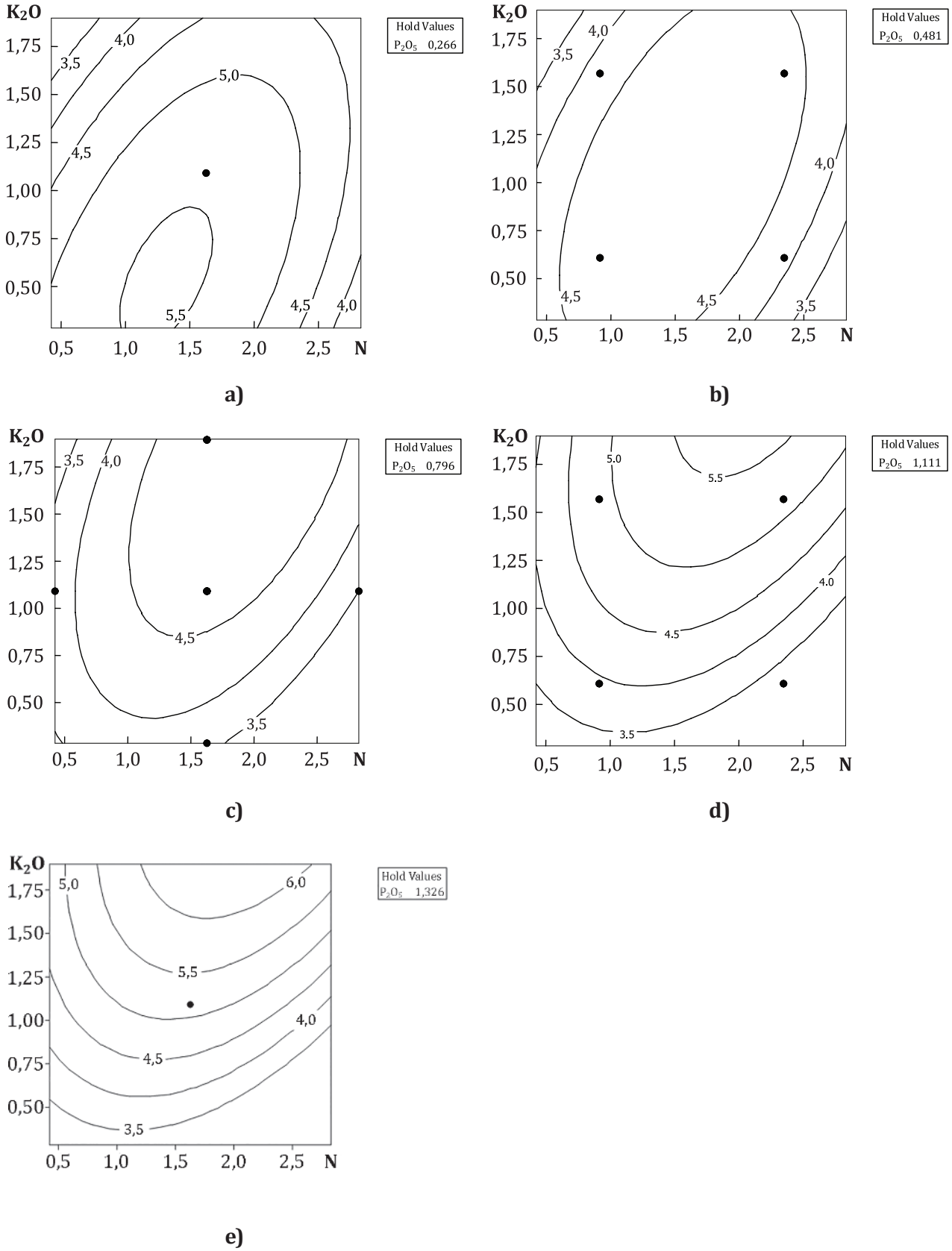
Table A.6 — Optimal settings and predicted yield given by Minitab's® response optimizer

Prediction for yield **Multiple response prediction**				
Variable	Setting			
N	1,933 04			
P_2O_5	1,326			
K_2O	1,9			
Response	Fit	SE Fit	95 % CI	95 % PI
Yield	6,395	0,540	(5,191; 7,599)	(4,953; 7,837)

The estimated standard errors of the predicted response at a selection of settings have been calculated using the Prediction option in Minitab's® *Analyse response surface designs*. The result is shown in Table A.7. The setting suggested by the response optimizer which is in the fourth row of Table A.7 is substantially larger than estimated standard errors at the other settings in Table A.7 which are all inside the experimental region.

A.9.3 Application of contour plots to find optimal settings

With three predictor variables, a series of contour plots of the response surface in two variables for a range of appropriately chosen values of the third value are needed to explore the response surface.



NOTE The design points are shown as dots in the contour plots.

Figure A.3 — Contour plots of predicted responses for five fixed values of P_2O_5

The contour plots in Figure A.3 are slices cutting through the contours in three-dimensional space at the selected values of P_2O_5. The fixed values are selected as the values of P_2O_5 that are used in the experiment, see Table A.1. The design points are shown in the contour plots as dots. The highest expected responses are seen in the contour plots with the two extreme values of P_2O_5. But the single design points shown in those two plots are axial runs so moving away from this axial run in the slice means moving outside the design space. When interpreting regression models it is not recommended to extrapolate outside the region where one has data. So with this in mind the safe recommendation for an optimal setting would be to set N to a value between 1,5 and 1,7, set P_2O_5 to a value between 1,11 and 1,25, say, and set K_2O to a value close to 1,60 in actual units. Two example settings are given in the first two rows of Table A.7. The last column gives the distance in software coded units of the setting to the design centre. If this distance is larger than one, the setting is outside the experimental region.

Table A.7 — Settings of predictor variables and the corresponding estimated yield and the estimated standard error of the estimated yield. The first two rows are settings suggested by the inspection of contours in Figure A.3. For comparison, the centre point is given in the third row and the optimal setting suggested by the desirability function is given in the fourth row.

N	P_2O_5	K_2O	Estimated yield	Estimated standard error	Distance to design centre
1,692	1,184	1,512	5,50	0,236	0,900
1,699	1,233	1,540	5,67	0,276	0,996
1,629	0,796	1,089	4,69	0,145	0
1,933	1,326	1,900	6,40	0,540	1,437

A.9.4 Canonical analysis

Minitab® does not offer the results from a canonical analysis as part of the output from its Stat → DOE → Response Surface → Analyze Response Surface Designs menu, but it is possible to use facilities in Minitab® to make the calculations using the formulas in Annex E. The matrix $\boldsymbol{B}$ in Formula (E.8) can be constructed manually and read into a matrix in Minitab® and then the Calc → Matrices → Eigen Analysis... menu can be used to obtain the eigenvalues and the eigenvectors. If the column matrix $\boldsymbol{b}$ in Formula (E.8) is also made available in Minitab®, the coordinates of the stationary point can be obtained using Formula (E.9).and the facilities for matrix manipulations in the Calc → Matrices → Arithmetic... menu.

The standard result from the canonical analysis is given in Tables A.8 and A.9. Table A.8 gives the coordinates of the stationary point in software coded units and in original units as well as the predicted response at the stationary point.

The software coded coordinates are particularly useful for judging whether the stationary point is inside or outside the experimental region. In this case, the stationary point is well within the experimental region. But if one or more coordinates approach 1 or –1, it may be helpful to calculate the distance from the stationary point to the centre point of the design. In this case, it is

$$D_S = \sqrt{0,107^2 + (-0,264)^2 + 0,438^2} = 0,52 \quad \text{(A.5)}$$

Thus, the stationary point is well within the experimental region because points with a distance smaller than 1 measured in software coded units are considered to be inside the experimental region.

Table A.8 — Canonical analysis of response surface: Stationary point

	Stationary point		
	N	**P_2O_5**	**K_2O**
Software coded units	0,107	−0,264	0,438
Original units	1,758	0,656	1,444
Predicted value at the stationary point 4,835			

Table A.9 contains information that characterizes the response surface. The first column gives the eigenvalues and two are negative and one is positive, which means that the stationary point is a saddle point or a minimax. The eigenvector corresponding to the positive eigenvalue of 0,811 is given in the same row. The labels **N**, **P2O5** and **K2O** have been used for the coordinates even if the coordinates are in software coded units. The interpretation of the eigenvector is that, if moving away from the stationary point in increments proportional to the eigenvector, then, the response increases because the eigenvalue is positive. Similarly, if moving away from the stationary point in increments proportional to the other two eigenvectors, then, the response decreases because the eigenvalues are negative.

Table A.9 — Canonical analysis of response surface based on software coded variables: Eigenvalues and eigenvectors

Eigenvalues	**Eigenvectors**		
	N	**P2O5**	**K2O**
0,811	0,057	0,919	0,389
−0,458	0,513	−0,361	0,779
−1,299	0,856	0,155	−0,492
Stationary point is a saddle point			

So there is no single maximum inside the experimental region, and the optimal setting should be found towards the boundary of the experimental region. Since there is only one positive eigenvalue, this optimum shall be found in the direction of the eigenvector corresponding to the positive eigenvalue. Note, from the coordinates of this eigenvector in Table A.9, that the dominating coordinate is the one that corresponds to P_2O_5, so moving in the negative direction of this eigenvector largely corresponds to decreasing the P_2O_5 component below the value in the stationary point, which is 0,66 in original units. Similarly, moving in the positive direction of this eigenvector corresponds to increasing the P_2O_5 component above 0,66 in original units.

The findings here are reflected in the contour plots of Figure A.3. Note that P_2O_5 = 0,66 is between the second and the third contour plot. Comparing the first two contour plots with P_2O_5 = 0,27 and P_2O_5 = 0,48 it is clear that the response increases when P_2O_5 decreases. Similarly, comparing the last three contour plots where P_2O_5 = 0,8, P_2O_5 = 1,11 and P_2O_5 = 1,33 it is clear how the response increases with P_2O_5.

A.9.5 Conclusion

The analysis has concentrated on finding settings of the predictor variables that would maximize the expected response while keeping the predictor variables within the experimental region. This is a problem that is often solved with response surface methods. The recommendations can be evaluated based on Table A.6. The nominal settings that were used before the experiment are in the third row of Table A.6. Suggested settings are in rows one and two. The increase in expected yield is about 0,9 kg/plot, but this is at the cost of increasing the amount of fertilizer with almost the same amount. Fertilizers are expensive so this is hardly worthwhile.

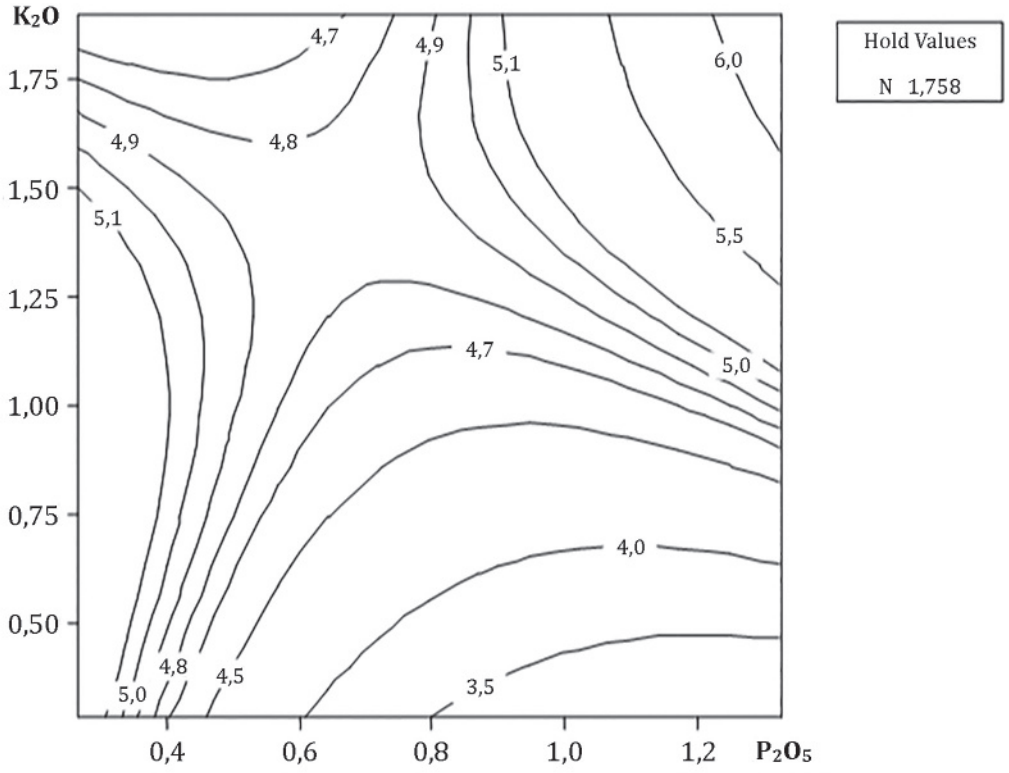
But it is possible to answer another question with these data: Is it possible to increase the yield to 5 kg/plot or above using different combinations of the three fertilizer components and reducing the total amount of fertilizer? This question can be answered using information from the canonical analysis and illustrated by an appropriate contour plot and this approach will be considered first. Alternatively,

the desirability function in Minitab's® Contour Profiler can be used to guide to a solution. This approach is explained in connection with Figure A.6.

Recall that moving away from the stationary point in the opposite direction of the eigenvector corresponding to the positive eigenvalue would also increase the response. This is the eigenvector in the first row of Table A.9. Noting that the coefficient corresponding to N is very small, so the amount of N changes very slowly, the options are best described by a contour plot where N is held fixed at the value at the stationary point, i.e. N = 1,758. This contour plot is shown in Figure A.4. It shows how the yield increases most rapidly when moving away from the stationary point in increments proportional to (P_2O_5, K_2O) = (0,919, 0,389). The solutions offered in Table A.6 were sought in the positive direction corresponding increasing the P_2O_5 and the K_2O components, i.e. in the upper right hand corner. Now, focus is on decreasing the P_2O_5 and the K_2O components while having an estimated yield of 5 kg/plot or above. It appears from the contour plot that quite a variety of settings of P_2O_5 and K_2O in the left hand side of the contour plot might achieve this goal. These settings, however, are close to the boundary of the experimental region where the conclusions are less certain, as is being emphasized by the increase in the estimated standard errors of the predictions. So, the recommendation would be to confirm the expected results at desired settings by more experimentation. This could be simply some confirmation runs, or a more elaborate design, such as a new response surface design, for example.

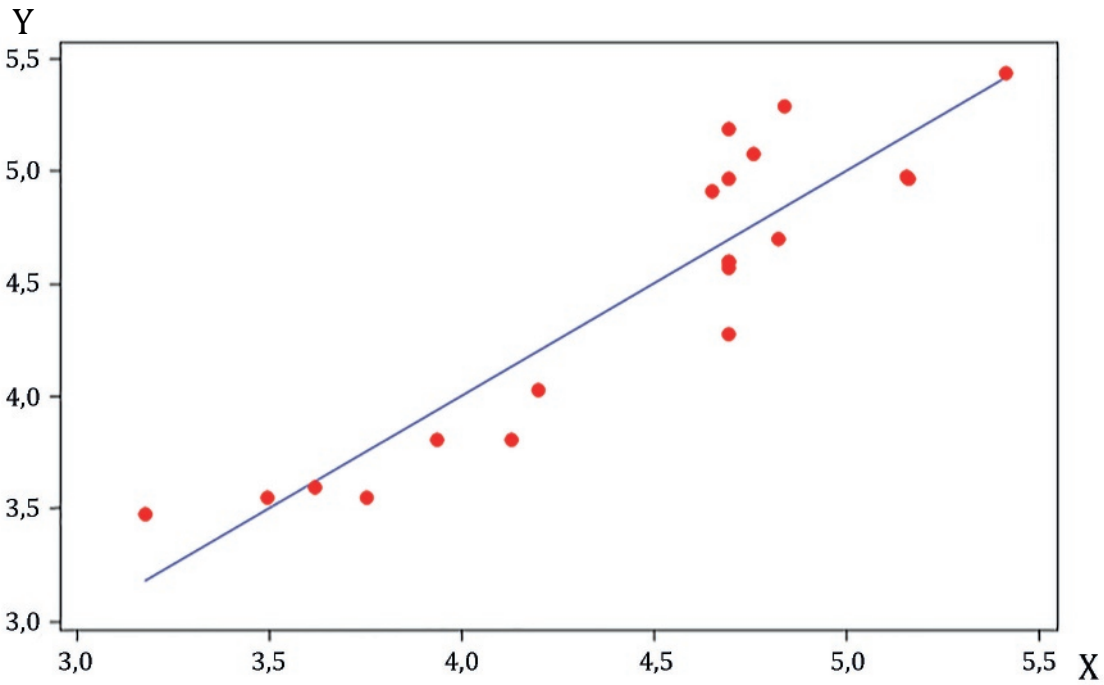
If a more elaborate design is chosen, it would be advisable to consider blocking in an attempt to reduce the random variation. Advice on blocking in response surface experimentation can be found in the References [2] and [7].

The scatterplot of observed yields against fitted values in Figure A.5 shows a rather large variation around the regression line showing that reducing the variation should be a major concern.



NOTE The left hand side of the plot shows an area of settings where smaller amounts of K_2O and P_2O_5 than at the nominal setting give a higher predicted yield than at the nominal settings.

Figure A.4 — Contour plot where N is held at 1,758



Key

X fitted values Y yield

NOTE For a perfect fit, the points would be on the blue line. The variation around the regression line is rather large. This plot is equivalent to the plot of residuals against fitted values in Figure A.1.

Figure A.5 — Scatterplot of observed yield against fitted values

Minitab's® Response Optimizer can assist in solving the problem of having a predicted yield of 5 kg/plot or above using combinations of the three fertilizer components and a reduction of the total amount of fertilizer used per plot. For this purpose the total amount of fertilizer (The variable Amount in Figure A.6) per plot is introduced as a response variable and a second order response model with predictors with N, P_2O_5 and K_2O is fitted. This is a mere formality to make Amount available in the Response Optimizer. In the setup window, it is chosen to minimize Amount with target 0 and to maximize Yield with lower limit set to 5, reflecting the desire to have response above five, and target arbitrarily set to seven. The result is seen in Figure A.6. It should be noticed that the suggested settings of both P_2O_5 and K_2O are at their extreme lower levels. So, just as was the case with the application of the desirability function leading to Figure A.2, the solution here is outside the experimental region. As always, when experimentation suggests that settings outside the experimental region might be optimal, such hypotheses should be confirmed with further experimentation.

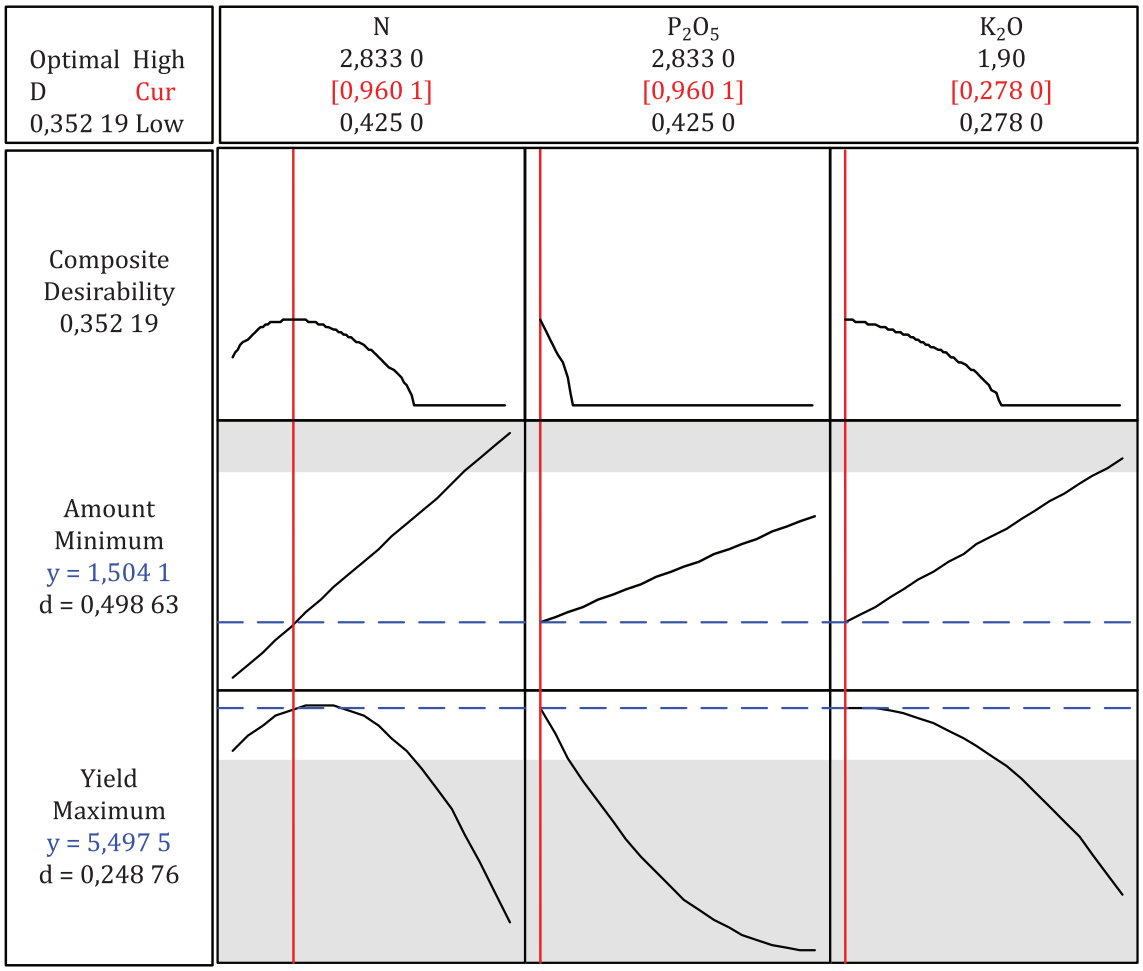


Figure A.6 — Optimization results based on the Desirability Function using two responses Yield and Total Amount of Fertilizer

Annex B
(informative)

Optimization of the button tactility using central composite design[3)]

B.1 Purpose of the experiment

Button tactility (perceptibility by touch) is a critical requirement for wireless products such as two way radios and cellular phones. Poor button tactility will lead to customer complaints and dissatisfaction. The main purpose of this experiment is to provide good button tactility for two way radios by optimizing specification settings of the existing component using statistical tools.

The objective of the experiment is to achieve good button tactility on the emergency button. The study is focused on the emergency button due to higher customer complaints associated with this button on previous products.

Initially, a full factorial experiment using four factors was conducted to optimize the button tactility. The analysis of the full factorial experiment indicated that two of these factors were significant. The analysis also showed the presence of curvature which suggest that quadratic terms may be needed in the model. A response surface experiment was planned to estimate the quadratic terms.

B.2 Response variable

Only one response variable was considered in this experiment: button tactility. The button tactility is defined as snap ratio (%) = [(F1–F2)/F1] × 100 %, where

F1 is *actuation force* (gf) and

F2 is *return force* (gf).

To measure F1 and F2, a mechanical plunger is used to depress the button and the force curve graph is plotted (Figure B.1). The unit gf is gram force (9,806 65 × 10^{-3} Newton). The first peak on the graph is the actuation force and the lowest point after the first peak is the return force.

B.3 Predictor variables

As mentioned in B.1 an initial full factorial experiment with four factors was conducted. The four factors in the initial experiment were the following:

— A. Duro hardness button;

— B. Actuation force of dome;

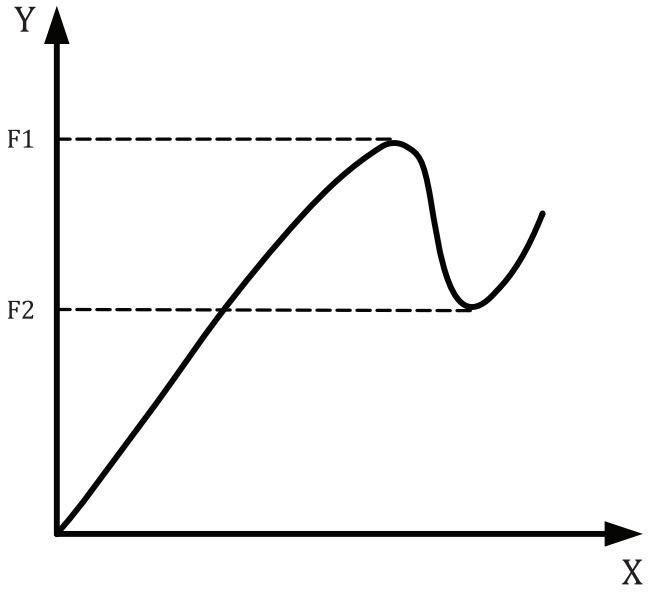
— C. Air vent width;

— D. Plunger length.

The initial experiment was conducted using the above four factors each at two levels. The analysis of this full factorial experiment showed that factor A and factor B had significant impact on the response variable button tactility, while factors C and D and interactions involving those factors were non-significant. The analysis also indicated the presence of interaction between A and B which suggested that quadratic terms needed to be included in the model.

3) Source: Harry Shah, USA.

Based on the results of the full factorial experiment, the two factors A, duro hardness button, and B, actuation force of dome, were selected for a response surface experiment. In the following, the shorthand notation DHB for duro hardness button and AFD for actuation force of dome will be used.



Key

X travel distance, in millimetres (mm)

Y force, in gram force (gf)

Figure B.1 — Button depress versus force curve

Table B.1 — Coded and actual levels of the experimental factors

Coded levels	–1,25	–1,00	0,00	+1,00	+1,25
Duro hardness button (DHB)	40	44	60	76	80
Actuation force of dome (AFD)	120	128	160	192	200
NOTE Coded levels are in the top row and actual levels are in the following rows.					

B.4 Identification and estimation of measurement systems

The repeatability and reproducibility of the measurement system for the associated response and predictor variables was deemed adequate for the objectives of this experiment, so no further work was initiated.

B.5 Selection of the settings of the predictor variables in the design

B.5.1 General considerations

Coded values of duro hardness button (DHB) and actuation force of dome (AFD) are denoted by x_1 and x_2, respectively, and the conversions from the actual levels given in Table B.1 to the coded levels are given as

$$x_1 = \frac{\text{Duro hardness button} - 60}{16} \tag{B.1}$$

$$x_2 = \frac{\text{Actuation force of dome} - 160}{32} \quad \text{(B.2)}$$

B.5.2 Factorial levels

The factorial levels are coded as −1 and +1, so the factorial levels in actual or natural units are given in the two columns labelled −1,00 and +1,00 in Table B.1. Factorial levels of the predictor variables in this experiment are the following:

— Duro hardness button (DHB): 44 and 76 (coded as −1 and 1 respectively);

— Actuation force of dome (AFD): 128 and 192 (coded as −1 and 1 respectively).

The factorial levels have been given their name because those levels are used to define a factorial design, which is an important building block in a CCD. The factorial design is typically a full two-level factorial design or a fractional two-level design if the number of factors is large.

B.5.3 Axial levels

These axial levels are the extreme values the predictor variables can have, so the axial levels in actual or natural units are given in the two columns labelled −1,25 and +1,625 in Table B.1. In this experiment, the axial levels are the following:

— Duro hardness button (DHB): 40 and 80 (coded as −1,25 and 1,25 respectively);

— Actuation force of dome (AFD): 120 and 200 (coded as −1,25 and 1,25 respectively).

B.5.4 Centre levels

The centre levels are the average of the two factorial levels, and it is also the average of the two axial levels. Its coded value is 0. In this experiment the centre levels were the nominal values of the predictor variables. But this is not a requirement. The experimenter is free to choose the region of experimentation where the response surface is being investigated. In this experiment, the centre levels are the following:

— Duro hardness button (DHB): 60 (coded as 0);

— Actuation Force of dome (AFD): 160 (coded as 0).

B.6 Experimental design

B.6.1 General

The levels of the individual predictor variables defined in B.5 are combined to define the runs of the experiment. The three types of runs in the CCD are the factorial, the star (or axial) and the centre runs.

B.6.2 Factorial runs

In the factorial runs, all predictor variables are at their factorial levels. The factorial runs usually constitute a full factorial or a fractional factorial design. In this experiment, a full factorial design is chosen, so the number of factorial runs is $2^2 = 4$ The factorial runs are given as serial numbers 1 to 4 in Table B.2. Their coded levels are given in columns 3 and 4 while the actual levels are given in columns 5 and 6.

B.6.3 Axial runs

In the axial runs, one of the predictor variables is at its upper or lower extreme level and the rest of the predictor values are at their centre value. The total number of axial runs in this experiment is 4. The axial runs are given as serial number 5 to 8 in Table B.2. Their coded levels are given in columns 3 and 4 while the actual levels are given in columns 5 and 6.

B.6.4 Centre runs

In the centre runs, all the predictor variables are at their centre level. The centre runs were replicated three times in this experiment. The three centre runs are given as serial number 9 to 11 in Table B.2. Their coded levels are given in columns 3 and 4 while the actual levels are given in columns 5 and 6.

B.7 Data generated by the experiment

The response variable is measured once for each setting of the predictor variables. All the information about the predictor variables and the response is collected in Table B.2. The first column gives the order in which the runs were made in the experiment. The second column gives a so-called serial number of the runs which can be used to identify the type of run. The serial number of a run is the number in the serial order. In this experiment the factorial runs have serial numbers 1 to 4, followed by the axial runs which have serial numbers 5 to 8, and finally the three centre runs have serial numbers 9 to 11. The runs are given in coded levels in columns 3 and 4 and in actual levels in columns 5 and 6.

Table B.2 — Design and data collected during the experiment

Run order	Serial number	x_1	x_2	Duro hardness button (DHB)	Actuation force of dome (AFD)	Button tactility, Y
1	11	0,00	0,00	60	160	31,86
2	1	–1,00	–1,00	44	128	21,15
3	4	1,00	1,00	76	192	33,07
4	3	–1,00	1,00	44	192	24,72
5	2	1,00	–1,00	76	128	30,39
6	10	0,00	0,00	60	160	31,09
7	6	1,25	0,00	80	160	26,74
8	7	0,00	–1,25	60	120	26,67
9	5	–1,25	0,00	40	160	14,79
10	8	0,00	1,25	60	200	33,84
11	9	0,00	0,00	60	160	33,08

B.8 Analysis of results

B.8.1 Software used for the analysis

The data are analysed with the rsreg procedure in SAS® 9.4.

B.8.2 Transfer function

The estimated transfer function (the estimated systematic part of the model) for Button Tactility is given by Formula (B.3), using actual or natural units

$$\begin{aligned}\text{Button Tactility} = & -74{,}848\,895 + 3{,}069\,693 * \text{DHB} + 0{,}004\,684 * \text{AFD} - 0{,}000\,435\left(\text{DHB} * \text{AFD}\right) \\ & - 0{,}022\,624 * \left(\text{DHB} * \text{DHB}\right) + 0{,}000\,275 * \left(\text{AFD} * \text{AFD}\right)\end{aligned} \quad \text{(B.3)}$$

The coefficients of the transfer function in Formula (B.3) are found in Table B.3. The values of the predictor variables in actual units are rather large and so the estimated coefficient may become small.

This is one argument for using coded units. It is good practice to perform the analysis in terms of the coded or software coded variables. The transfer function in coded units is as follows.

$$\begin{aligned}\text{Button Tactility} = {} & 31{,}514\,229 + 4{,}565\,263 x_1 + 2{,}135\,088 x_2 - 0{,}222\,500 * (x_1 * x_2) \\ & - 5{,}791\,643 (x_1 * x_1) + 0{,}281\,957 (x_2 * x_2)\end{aligned} \tag{B.4}$$

The coefficients of the transfer function in Formula (B.4) are found in Table B.4.The analysis is often performed with coded variables and the results are subsequently interpreted in terms of the actual variables.

B.8.3 Estimation of coefficients

Table B.3 — Estimated regression coefficients using actual units

Parameter	DF	Estimate	Standard error	t Value	Pr > |t|	Parameter estimate from Software Coded Data
Intercept	1	−74,848 895	38,487 109	−1,94	0,109 4	31,514 229
AFD	1	0,004 684	0,364 119	0,01	0,990 2	2,668 860
DHB	1	3,069 693	0,617 764	4,97	0,004 2	5,706 579
AFD*AFD	1	0,000 275	0,001 060	0,26	0,805 4	0,440 557
DHB*AFD	1	−0,000 435	0,002 167	−0,20	0,848 9	−0,347 656
DHB*DHB	1	−0,022 624	0,004 239	−5,34	0,003 1	−9,049 443

Table B.4 — Estimated regression coefficients using coded units

Parameter	DF	Estimate	Standard error	t Value	Pr > |t|	Parameter estimate from Software Coded Data
Intercept	1	31,514 229	1,256 212	25,09	<0,000 1	31,514 229
x2	1	2,135 088	0,831 129	2,57	0,050 1	2,668 860
x1	1	4,565 263	0,831 129	5,49	0,002 7	5,706 579
x2*x2	1	0,281 957	1,085 222	0,26	0,805 4	0,440 557
x1*x2	1	−0,222 500	1,109 254	−0,20	0,848 9	−0,347 656
x1*x1	1	−5,791 643	1,085 222	−5,34	0,003 1	−9,049 443

The analysis has been made using coded units, but for easy reference with the transfer function in actual units in in Formula (B.3) in B.8.2 the estimated regression coefficients for actual units are given in Table B.3. The rsreg procedure in SAS® presents some of the results in software coded units, but uses the term *Coded Data*. In order to avoid confusion, the term has been changed to *Software Coded Data* in the tables. Apart from this, the tables and figures in B.7 and B.8 have been produced by the rsreg procedure. Note that the coded variables x_1 and x_2 are called x1 and x2, respectively, in the tables.

Table B.5 shows the F-tests for the hypotheses that each predictor variable is not needed in the model. The test in the row labelled x2 is the test for the hypothesis that $\beta_2 = \beta_{12} = \beta_{22} = 0$, i.e. the hypothesis that the coefficients of all the terms in the model that involve x_2 are 0. These tests are the same whether coded variables, software coded variable or actual variables are used in the analysis. These tests should be used with caution. In this case, the hypothesis that x_2 is not needed in the model is not rejected, the p-value being 0,202 0, but inspection of individual p-values in Table B.4 indicates that the hypothesis $\beta_2 = 0$ should probably be rejected. The p-value is 0,050 1 in Table B.4 which is borderline.

Table B.5 however does give the information that most of the variation in the data are explained by x_1 (DHB).

Table B.5 — Are all predictor variables needed?

Factor	DF	Sum of squares	Mean square	F-value	Pr > F
x2	3	33,010 284	11,003 428	2,24	0,202 0
x1	3	288,875 252	96,291 751	19,56	0,003 4

Tables B.6 and B.7 are the ANOVA tables for the central composite design and they are often shown together as a single table, and indeed some of the entries rely on information from both tables. The F-value, 18,39, in the Linear row of Table B.6 is calculated using Formula (B.5).

$$\frac{\left(\frac{180{,}976\,619}{2}\right)}{4{,}921\,775} = 18{,}39 \qquad \text{(B.5)}$$

Here, the numbers in the numerator are found in the Sum of Squares and DF columns of Table B.6 and the denominator is the Total Error Mean Square from Table B.7.

Table B.6 is used to decide on the complexity of the model for the response surface. The most important question is whether a quadratic surface is needed. This question is answered by the F-value and the corresponding P value (Pr > F) in the Quadratic line of Table B.6. The P value is much smaller than the conventional significance level of 0,05. The conclusion is that a quadratic surface, or a second order model, is needed.

Table B.6 — ANOVA table for CCD for deciding the type of response surface

`	DF	Sum of squares	R-Square	F-value	Pr > F
Linear	2	180,976 619	0,515 2	18,39	0,005 0
Quadratic	2	145,501 951	0,414 2	14,78	0,008 0
Crossproduct	1	0,198 025	0,000 6	0,04	0,848 9
Total model	5	326,676 595	0,929 9	13,27	0,006 5

Table B.7 gives information on the fit of the second order model. The F-value compares the Lack of Fit estimate of the error to the Pure Error estimate of the error which is this design comes from the three replicates at the centre point. The test does not question the model.

Table B.7 — ANOVA table for CCD for lack of fit

Residual	DF	Sum of squares	Mean square	F-value	Pr > F
Lack of Fit	3	22,595 077	7,531 692	7,48	0,120 2
Pure error	2	2,013 800	1,006 900		
Total error	5	24,608 877	4,921 775		

B.9 Presentation of results — Optimization

B.9.1 General

Just looking at the transfer function (the estimated systematic part of the model), whether it is shown in actual units as in Formula (B.3) or in coded units as in Formula (B.4), does not reveal the properties of the response surface. It is important to realize that the second order polynomial in two or more variables can represent a variety of surfaces. The stationary point can be a maximum, a minimum or a saddle point. Furthermore, the stationary point can be located outside as well as inside the experimental region. The situation is only simple, if the object is to maximize the response and if the stationary point is a maximum located inside the experimental region, or, alternatively, if the object is to minimize the response and if the stationary point is a minimum located inside the experimental region, and then

the optimal settings is the stationary point. In all other cases the optimal settings needs to be found away from the stationary point but within the boundary of the experimental region. If the number of predictor variables is two, then contour plots can be very useful as illustrated in B.9.3.

B.9.2 presents the canonical analysis of the fitted second order response surface. Canonical analysis has been the basic tool to gain a precise understanding of the response surface. Although modern computer software has various impressive graphical and analytical tools, the knowledge obtained from a canonical analysis helps to fully exploit the graphical tools. The results of the canonical analysis in this experiment are that the stationary point is a saddle point that is located far outside the experimental region.

B.9.2 Canonical analysis

The standard result from the canonical analysis is given is Tables B.8 and B.9. Table B.8 gives the coordinates of the stationary point, $\boldsymbol{x}_S$, in coded units and in original units as well as the predicted response at the stationary point. Mathematical details of canonical analysis can be found in E.3.

Table B.8 — Canonical analysis of response surface based on coded variables: Stationary point

	Critical value	
Factor	Software coded	Uncoded
x1	0,370 673	0,463 341
x2	−2,882 706	−3,603 382
Predicted value at stationary point: 28,725 098		

As mentioned in B.8.3, the rsreg procedure in SAS® presents some of the results in software coded units, but uses the term *Coded Data*. Table B.8 has been taken from the SAS® output, but the column labelled Software Coded was originally labelled Coded in the SAS® output. The rsreg procedure gives some of the calculations in software coded units, which SAS® simply refer to as coded.

Table B.9 contains information that characterizes the estimated response surface. The first column gives the eigenvalues and they have different signs, so the stationary point is a saddle point. The eigenvector corresponding to the eigenvalue 0,443 740, for example, is given in the same row.

Table B.9 — Canonical analysis of response surface based on coded variables: Eigenvalues and eigenvectors

	Eigenvectors	
Eigenvalues	x1	x2
0,443 740	−0,018 308	0,999 832
−9,052 626	0,999 832	0,018 308
Stationary point is a saddle point.		

It is worth noting that the eigenvalues given in Table B.9 by the SAS® procedure rsreg are calculated based on software coded variables, even if the predictor variables that are used in the analysis are the coded variables x_1 and x_2 from Table B.2.

The distance of the stationary point to the design centre in coded units is

$$D_S = \sqrt{0,463\,341^2 + (-3,603\,382)^2} = 3,63 \tag{B.6}$$

So the stationary point is in this case outside the experimental region because the extreme points of the design, the factorial points and the axial points in this design are at a distance from the design centre of $\sqrt{2} = 1,41$ and $\alpha = 1,25$, respectively.

It may be useful to refer to the contour plot of Figure B.2 in the discussion of the canonical analysis because it displays the design points. The stationary point is $\boldsymbol{x}_S = (x_{S1}, x_{S2}) = (0,463\,341, -3,603\,382)$ so it is obvious that the stationary point is outside the experimental region. The eigenvector corresponding to the numerically largest eigenvalue, –9,052 626, is $\boldsymbol{e}_1 = \left(0,999\,832, 0,018\,308\right)$, which is close to (1,0) and the eigenvector corresponding to the numerically smallest eigenvalue, 0,443 740, is $\boldsymbol{e}_2 = (-0,018\,308,\ 0,999\,832)$ which is close to (0,1).

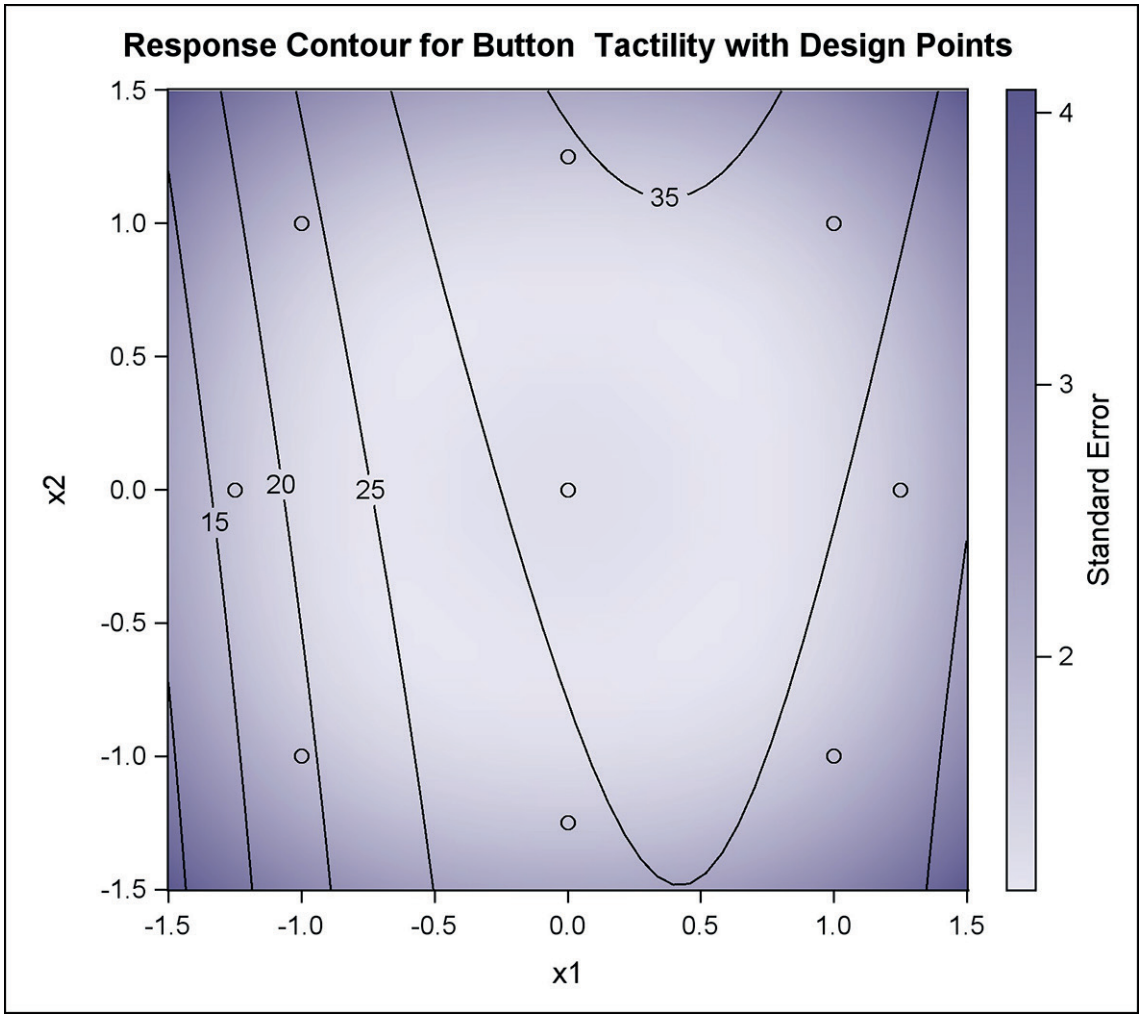
Moving away from the stationary point in increments proportional to $\boldsymbol{e}_1$ will decrease the predicted response, and moving away from the stationary point in increments proportional to $\boldsymbol{e}_2$ will increase the predicted response. This means that in the design region the fitted response surface is a rising ridge. The rising ridge is a straight line through the stationary point and pointing in the direction of $\boldsymbol{e}_2$. The formula of the rising ridge is as follows:

$$x_2 = -54{,}612\, x_1 + 21{,}7 \tag{B.7}$$

The ridge intersects the line x_2 = –1,25 (AFD = 120) at x_1 = 0,420 24 (DHB = 66,724) and the line x_2 = 1,25 (AFD = 200) at x_1 = 0,374 46 (DHB = 65,991). In the experimental region the ridge is very well approximated by the vertical line x_1 = 0,4 (DHB = 66,4).

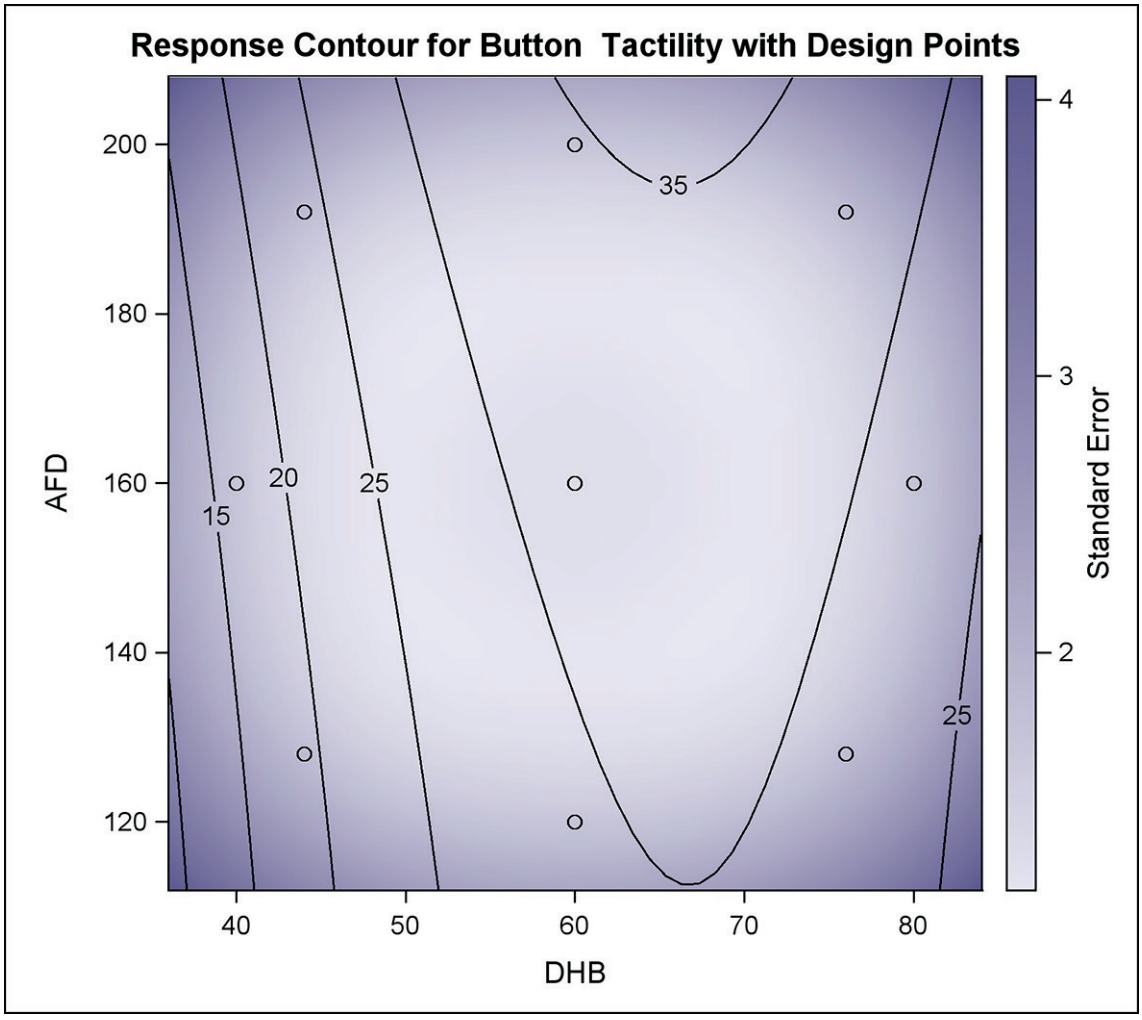
B.9.3 Application of contour plots to find optimal settings

With only two predictor variables most conclusions can be drawn from a contour plot as in Figure B.2 where coded values are used on the axes or in Figure B.3 where actual values are used on the axes. A high value of button tactility is desired, and it is obvious from the contour plots that a predicted response above 35 can be achieved with x_1 around 0,3 (DHB around 65) and x_2 around 1,25 (AFD close to 200).



NOTE The shading is used to indicate the estimated standard error of the predicted response, see the colour code on the right hand side of the figure.

Figure B.2 — Contours of the estimated response surface for button tactility with the design points indicated by open circles



NOTE Figure B.3 is the same as Figure B.2 except that Figure B.3 uses the actual values of the predictor variables on the axes.

Figure B.3 — Contours of the estimated response surface for button tactility with the design points indicated by open circles

B.9.4 Application of ridge analysis

With only two predictor variables, there is really no need to apply other techniques than contour plots or canonical analysis, but it may be useful to present a technique that is potentially useful with a large number of predictor variables. The method of ridge analysis was introduced by Hoerl in 1959, but a good introduction is given in Reference [8]. Mathematical details are given by Draper in Reference [9].

The SAS® procedure rsreg has implemented ridge analysis, which can be used to find a ridge of optimal responses. This ridge is not necessarily the same as the one found using canonical analysis. But it does give roughly the same answer in terms of optimal settings. The ridge starts at a given point $\mathbf{x}_0$ which is chosen well within the experimental region, and in this example the design centre is chosen as the starting point $\mathbf{x}_0$. The estimated response surface is then investigated on a series of circles centred at $\mathbf{x}_0$ and on each circle the setting that gives the optimal response is found. The ridge is the series of optimal settings.

The ridge analysis can be summarized both in tabular form as in Table B.10 and in graphical form as in Figure B.3. The Coded Radius column of Table B.10 gives the radius of the circle in software coded units. It makes good sense to use software coded units to calculate the radius; for in software coded units the circle with radius 1 can be considered to be the boundary of the experimental region. The

Estimated Response in the second column is the maximal response on the circle with the given radius. The Standard Error in the third column is the estimated standard error of the maximal estimated response. The standard error increases as the radius increases. Finally, the last two columns give the settings that correspond to the maximal estimated responses in column two. The settings are labelled "Uncoded Factor values" in Table B.10 but in the SAS® language this means the variables supplied by the user as predictor variables, and in the analysis presented here the coded variables were used as predictor variables.

Table B.10 — Summary of the ridge analysis in tabular form

Estimated ridge of maximum response for variable button tactility				
			Uncoded factor values	
Coded radius	Estimated response	Standard error	x2	x1
0,0	31,514 229	1,256 212	0	0
0,1	32,074 250	1,249 734	0,067 751	0,105 047
0,2	32,520 660	1,231 254	0,168 801	0,184 408
0,3	32,899 982	1,203 886	0,293 230	0,233 754
0,4	33,248 325	1,173 959	0,425 520	0,262 551
0,5	33,584 685	1,151 747	0,558 739	0,280 063
0,6	33,918 115	1,151 327	0,691 107	0,291 326
0,7	34,253 166	1,188 869	0,822 374	0,298 875
0,8	34,592 327	1,278 712	0,952 654	0,304 057
0,9	34,937 067	1,429 173	1,082 120	0,307 640
1,0	35,288 304	1,641 494	1,210 926	0,310 092

Figure B.4 is basically a summary of most of the information in Table B.10. The top panel in Figure B.4 shows how the maximal estimated response in this case increases with the radius. The second panel shows the optimal settings as a function of the radius. Here software coded units are used which makes sense, because the software coded units always range from 0 to 1.

B.9.5 Conclusion

All the investigations of the estimated response surface in B.9.2 to B.9.4 conclude that the maximum value of button tactility is obtained when DHB is around 65 and AFD is around 200. AFD at 200 is the upper boundary of the experimental region, so higher values are not recommended without further investigation. Standard Duro Hardness is available in increments of 5, so the team selected a Duro Hardness Button of 65. Then the potential optimal values for DHB and AFD were as follows:

— Duro Hardness Button = 65

— Actuation Force of Dome = 200

This setting is close to the ridge of maximum response given in Table B.10. The point in the bottom line of Table B.10 corresponds to the actual values DHB $= 16x_1 + 60 = 64,96$ and AFD $= 32x_2 + 160 = 198,75$, which are indeed close to the chosen potential optimal values of DHB=65 and AFD=200.

Table B.11 — Estimated response and standard error at the potential optimal values DHB = 65 and AFD = 200

Estimated response	Standard error	Uncoded factor values	
		DHB	AFD
35,397 786	1,718 440	65	200

The estimated response at the potential optimal setting is $\hat{y} = 35,397786$ with estimated standard error 1,718 440 as can be seen from Table B.11. Using these values, the 95 % confidence interval for the predicted response at the chosen setting can be calculated as

$$\hat{y} \pm t_{.975}(v) \times \text{std.err} = 35,397\,786 \pm 2,570\,6 \times 1,718\,440 = [30,98, 39,82].$$

This is a fairly wide confidence interval, and there are a number of reasons for that. First, there are only $v = 5$ degrees of freedom, so the t-fractile is relatively large. Secondly, the chosen setting is at the boundary of the experimental region, so the standard error is larger than in the centre of the design.

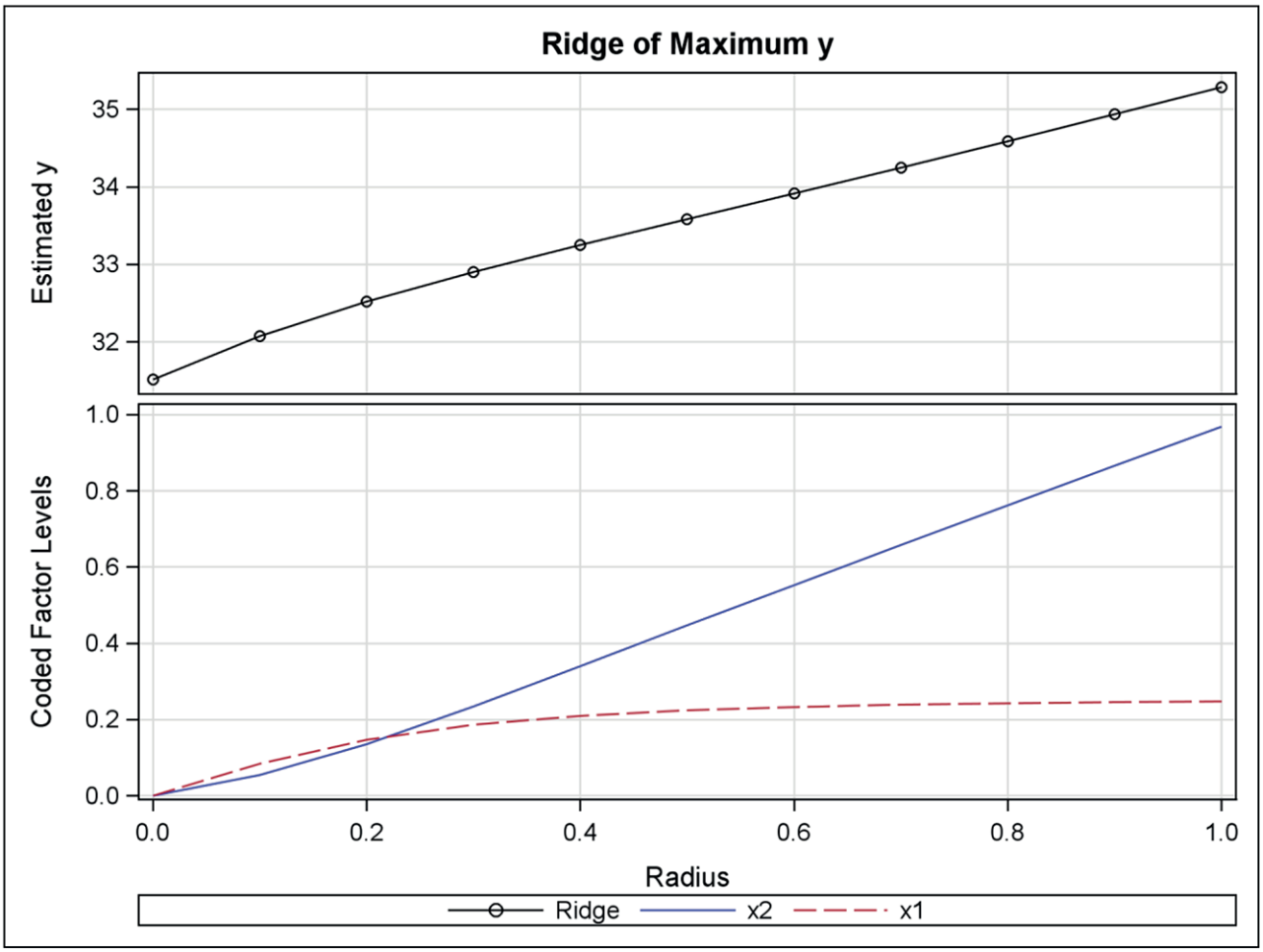


Figure B.4 — Graphical summary of the ridge analysis

B.10 Confirmation

To validate the model, five confirmation test runs were conducted at the optimum settings. The Button Tactility was measured for these five runs. The average Button Tactility for the five runs was found to be 31,76 %.

Since the measured average value of Button Tactility (31,76 %) for the confirmation runs falls within the 95 % confidence interval of the predicted value, the team concluded that the response surface model adequately predicts future observations.

Actually, the deviation from the average of 5 observations and the predicted response could have been even bigger than the one seen here. The reason is that the relevant interval to consider is not the confidence interval of the mean, but the 95 % prediction interval of an average of 5 independent

observations. This prediction interval is wider than the confidence interval of the mean because it takes the variability of the new observations into account. This prediction interval is calculated as

$$\hat{y} \pm t_{.975}(v) \times \sqrt{\frac{\text{std.dev}^2}{5} + \text{std.err}^2} = 35{,}397\,786 \pm 2{,}570\,6 \times \sqrt{\frac{4{,}921\,775}{5} + 1{,}718\,440^2} = [30{,}30,\ 40{,}50].$$

Here, the squared standard deviation is found as Total error mean square in Table B.7 and the standard error is found in Table B.11.

Annex C
(informative)

Semiconductor die deposition process optimization[4)]

C.1 Purpose of the experiment

Semiconductor dies (chips) form an integral part of electronic devices such as satellites and cell phones. Numerous dies are manufactured simultaneously on a thin round silicon wafer.

In one of the manufacturing steps, a silicon dioxide (SiO_2) layer is deposited over an aluminium layer to create an inter-level dielectric (insulator) between metals. The metals are patterned into strips (runners) that carry electrical current. The oxide layer is deposited using a PETEOS (Plasma Enhanced TetraEthylOrthoSilicate) process.

During the PETEOS deposition process, a wafer is heated in a chamber and gases are introduced through a shower head. A radio (high) frequency power strips electrons from TEOS molecules, ionizing them as they pass through a shower head and creating a plasma. The TEOS decomposes depositing a silicon dioxide layer on the wafer. A designed experiment was performed to

— keep the oxide deposition rate above 127 Ångstroms per minute,

— minimize the oxide deposition rate Non-Uniformity, and

— take stress to a target value of $1{,}5 \times 10^9$ dynes/cm^2.

C.2 Response variable

Three experimental responses were considered.

a) Average oxide **Deposition Rate.**

b) Deposition rate **Non-Uniformity.**

c) **Stress**.

The oxide deposition rate (deposited thickness divided by deposition time) was measured at 13 sites on each experimental wafer, and the average $\bar{x}$ and standard deviation s were calculated. The average oxide deposition rate $\bar{x}$, which reflects throughput, should be maximized.

Non-Uniformity (or coefficient of variation), given by $\frac{s}{\bar{x}}100\%$, expresses the standard deviation as a percentage of the mean. Non-Uniformity should be minimized to reduce differences in oxide thickness across the wafer.

The underlying aluminium layer on the wafer has a higher expansion coefficient than the glass layer of silicon dioxide that is deposited above it. If the aluminium layer contracts with respect to the SiO_2 layer (tensile force) the wafer will become convex and the glass might crack, while if the aluminium layer expands relative to the SiO_2 layer (compressive force) the wafer will become concave (bow) which might open up the aluminium leads. Stress measures the amount of pressure on a wafer resulting from all layers. It is desirable for stress to have a small positive value (slight wafer convexity).

4) Source: Veronica Czitrom, Statistical Training & Consulting, USA/Singapore.

C.3 Predictor variables

Two experimental factors were considered.

a) **Pressure** inside the chamber.

b) **Spacing** between the wafer and the shower head.

The range of the factor levels was restricted.

Pressure could only be varied between 8 torr and 9 torr, because pressure below 8 torr led to plasma arcing, and pressure above 9 torr led to high particle count due to gas phase reactions.

The process was expected to improve for higher spacing and worsen for lower spacing. The low level of Spacing was set at 180 mils (current operating conditions, where one mil is 1/1 000 of an inch), and the high level at 200 mils.

C.4 Identification and estimation of measurement systems

The repeatability and reproducibility of the measurement system for the associated response and predictor variables was deemed adequate for the objectives of this experiment, so no further work was initiated.

C.5 Selection of the settings of the predictor variables in the design

C.5.1 General considerations

Previous experimental results indicated curvature in Non-Uniformity. A central composite experimental design was selected to estimate curvature and optimize the process. Due to the restrictions on the factor levels, a face-centred design was chosen. A face-centred design requires only three levels of each factor and this may be an advantage when the range of factors is restricted.

Coded values of pressure and spacing are denoted by x_1 and x_2, respectively, and the conversions from the actual levels given in Table C.1 to the coded levels are given as

$$x_1 = \frac{\text{Pressure} - 8,5}{0,5},$$
$$x_2 = \frac{\text{Spacing} - 190}{10}. \qquad \text{(C.1)}$$

C.5.2 Factorial levels

The factorial levels in this experiment were chosen as the extreme levels possible, see C.3.

The factorial levels are coded as −1 and +1, so the factorial levels in actual or natural units are given in the two columns labelled −1 and +1 in Table C.1. Factorial levels of the predictor variables in this experiment are the following:

— Pressure (A): 8 and 9 (coded as −1 and 1 respectively);

— Spacing (B): 180 and 200 (coded as −1 and 1 respectively).

The factorial levels have been given their name because those levels are used to define a factorial design, which is an important building block in a CCD.

C.5.3 Axial levels

In a face-centred design, the axial levels are the same as the factorial levels.

C.5.4 Centre levels

The centre levels are the average of the two factorial levels, and they are also the average of the two axial levels. Its coded value is 0.

— Pressure (A): 8,5 (coded as 0)

— Spacing (B): 190 (coded 0)

In this experiment, the centre levels were not the nominal values of the predictor variables. It was undesirable to centre the design on the current operating conditions, because the process was expected to worsen for spacing below 180.

Table C.1 — Coded and actual levels of the experimental factors

Coded levels	–1	0	+1
Pressure (A)	8	8,5	9
Spacing (B)	180	190	200

NOTE Coded levels are in the top row and actual levels are in the following rows. This is a face-centred design so each factor has three levels.

C.6 Experimental design

C.6.1 General

The levels of the individual predictor variables defined in C.5 are combined to define the runs of the experiment. The three types of runs in the CCD are the factorial, the star (or axial) and the centre runs.

C.6.2 Factorial runs

The factorial design is typically a full two-level factorial design or a fractional two-level design if the number of factors is large. In this case with only two factors the factorial runs constitute a replicated full 2^2 design. The factorial runs are shown in actual levels as the first 8 runs in Table C.2.

Table C.2 — Face-centred central composite design in the two factors Spacing and Pressure

Run No.	Pattern	Pressure	Spacing	Run type
3	−−	8	180	Factorial run
13	−−	8	180	Factorial run
6	+−	9	180	Factorial run
10	+−	9	180	Factorial run
4	−+	8	200	Factorial run
12	−+	8	200	Factorial run
7	++	9	200	Factorial run
11	++	9	200	Factorial run
1	00	8,5	190	Centre run
8	00	8,5	190	Centre run
15	00	8,5	190	Centre run
9	0a	8,5	180	Axial run
2	a0	8	190	Axial run
5	A0	9	190	Axial run
14	0A	8,5	200	Axial run

C.6.3 Axial runs

In the axial runs, one of the predictor variables is at its upper or lower extreme level and the rest of the predictor values are at their centre value. The total number of axial runs in this experiment is 4. The axial runs are shown in actual levels as the last three runs in Table C.2. The axial runs are not replicated. Note that the axial levels are the same as the factorial levels showing that this design is a face-centred CCD.

C.6.4 Centre runs

In the centre runs, all the predictor variables are at their centre level. The centre runs were replicated 3 times in this experiment. The 3 centre runs are given as Run no. 1, 8 and 15 in Table C.2. Their actual levels are given in columns 3 and 4.

C.7 Data generated by the experiment

The three response variables are measured once in each run. All the information about the predictor variables and the response is collected in Table C.3. The first four columns are the same as the first four columns of Table C.2, but in Table C.3 the runs are listed in the order in which they were made. The last three columns give the observations of the three response variables Deposition, Rate Non-Uniformity, and Stress.

Table C.3 — Experimental Results for two factors and three responses

Run no.	Pattern	Pressure	Spacing	Dep Rate	Non-Uniformity	Stress
1	00	8,5	190	128	1,01	1,56
2	a0	8	190	129	0,40	1,63
3	−−	8	180	131	1,22	1,68
4	−+	8	200	128	0,61	1,45
5	A0	9	190	127	2,63	1,50
6	+−	9	180	128	3,19	1,57
7	++	9	200	127	1,37	1,32
8	00	8,5	190	129	0,94	1,58
9	0a	8,5	180	130	1,89	1,65
10	+−	9	180	129	3,13	1,52
11	++	9	200	126	1,45	1,30
12	−+	8	200	129	0,74	1,47
13	−−	8	180	131	1,22	1,68
14	0A	8,5	200	127	0,50	1,43
15	00	8,5	190	128	0,97	1,58

C.8 Analysis of results

C.8.1 General

The essential feature of this experiment is that it investigates three responses where

— the oxide deposition rate must be above 127 Ångstroms per minute;

— the oxide deposition rate Non-Uniformity should be minimized;

— the stress must be set to a target value of $1,5 \times 10^9$ dynes/cm^2.

Second order models will be fitted to the three responses, but only the analysis for stress will be presented in detail. The fitted response surfaces will be investigated to see if and how the three criteria can be met simultaneously. The analysis will be performed with coded variables x_1 and x_2 as predictor variables.

C.8.2 Software used for the analysis

The experiment was analysed using JMP®, version 10, SAS Institute Inc., Cary, NC, USA.

C.8.3 Transfer function

The transfer function is called the prediction expression in JMP®. The coefficients of the transfer function are found from Table C.4.

$$\text{Stress} = 1,581 + -0,113 * x_1 - 0,070 * x_2 - 0,004 * \left(x_1 * x_2\right) - 0,052 * \left(x_1 * x_1\right) - 0,027 * \left(x_2 * x_2\right) \quad \text{(C.2)}$$

C.8.4 Estimation of coefficients

The table of parameter estimates for stress with x_1 and x_2 as predictor variables is Table C.4. Note that x_1 and x_2 are labelled as x1 and x2 in output copied from JMP®. All estimates are significant except for the coefficient of x_1x_2, but this term is nevertheless kept in the model.

Table C.4 — Parameter estimates

Term	Estimate	Std Error	t Ratio	Prob > |t|
Intercept	1,580 857 1	0,008 855	178,52	<0,000 1*
x1	−0,113	0,005 522	−20,46	<0,000 1*
x2	−0,07	0,005 522	−12,68	<0,000 1*
x1*x1	−0,052 143	0,010 436	−5,00	0,000 7*
x1*x2	−0,003 75	0,006 174	−0,61	0,558 6
x2*x2	−0,027 143	0,010 436	−2,60	0,028 7*

Table C.5 — Analysis of variance

Source	DF	Sum of squares	Mean square	F Ratio	Prob > F
Model	5	0,192 095 36	0,038 419	125,9806	<0,0001*
Error	9	0,002 744 64	0,000 305		
C. Total	14	0,194 840 00			

Table C.6 — Lack of fit

Source	DF	Sum of squares	Mean square	F Ratio	Prob > F
Lack of fit	3	0,000 827 98	0,000 276	0,864 0	0,509 2
Pure error	6	0,001 916 67	0,000 319		
Total error	9	0,002 744 64			

The analysis of variance table and the lack of fit tables are provided by JMP® and are copied to Table C.5 and Table C.6, respectively.

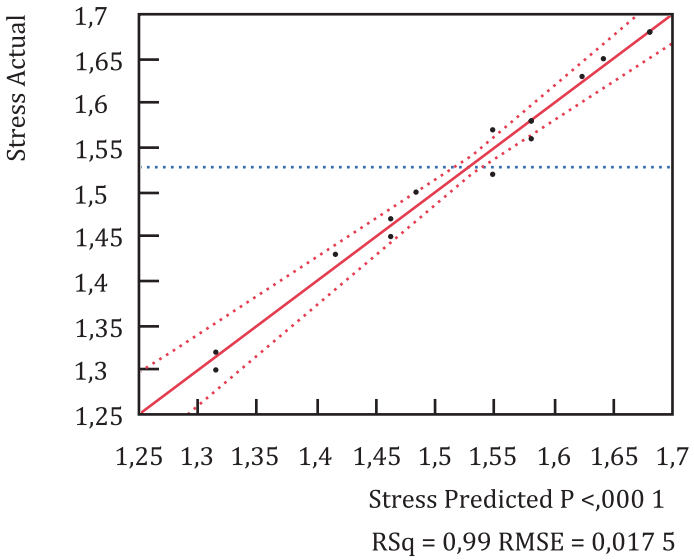
In Table C.5 the very large and strongly significant F-test is noted. In Table C.6 the Error Mean Square is divided into its Lack of Fit and Pure Error components. The F-test for lack of fit is calculated and the *p*-value is reported as 0,509 2. The model is not rejected.

Tests that each of the explanatory variables is not needed in the model are called Joint Factor tests in JMP®, and they are given in Table C.7. The row x1 gives the test of the hypothesis that $\beta_1 = \beta_{12} = \beta_{11} = 0$, and the row x2 gives the test for the hypothesis that $\beta_2 = \beta_{12} = \beta_{22} = 0$. The results are not surprising in view of the tests of the individual parameters given in Table C.4.

Table C.7 — Joint factor tests – Are all predictor variables needed?

Term	DF	Sum of squares	F Ratio	Prob > F
x1	3	0,135 415 36	148,014 2	<0,0001*
x2	3	0,051 175 36	55,936 6	<0,0001*

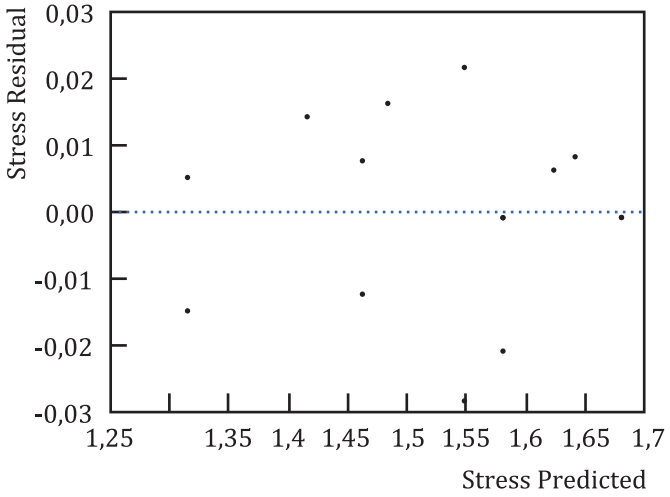
The test reported in Tables C.4 to C.7 can be supplemented by some of regression plots used for model checking in multivariate regression. Figure C.1 shows the *Actual by Predicted Plot which* shows the actual observations plotted against the predicted values from the model.



NOTE The observations are lying closely around the identity line showing a good fit of the model.

Figure C.1 — Actual observation plotted against predicted values

The residuals are plotted against fitted values in Figure C.2. The residuals are lying closely around the horizontal line through 0 indicating a satisfactory fit.



NOTE The residuals are lying symmetrically around the horizontal line through 0 indicating a satisfactory fit of the model.

Figure C.2 — Residuals plotted against predicted value

C.9 Presentation of results — Optimization

C.9.1 General

This example has three responses. In C.9.2, the fitted response surfaces are characterized using canonical analysis. In C.9.3 and C.9.4, the focus is on finding the settings where the different requirements of the three responses are met. These requirements are mentioned in C.1. C.9.3 illustrates the desirability function and C.8.4 illustrates the Contour Profiler.

C.9.2 Canonical analysis

JMP® also gives the results of the canonical analysis under the headline *Response surface*. The stationary point is called *Solution* and is given in Table C.8. The stationary point is characterized as a maximum although this observation is based on the eigenvalues, which are given in Table C.9.

Table C.8 — Solution for stress

Variable	Critical value
x1	−1,039776
x2	−1,217647
Solution is a Maximum	
Critical values outside data range	
Predicted Value at Solution 1,682 222 2	

The critical value (stationary point) is given in coded units in Table C.8, but in actual units the critical value is Spacing = 179,602 2 and Pressure = 7,891 2. The critical value for stress is just outside the lower left hand corner of the experimental region, or the data range, as explained in the JMP® output in Table C.8.

Table C.9 — Canonical curvature

	Eigenvalues and Eigenvectors	
Eigenvalue	−0,027 0	−0,052 3
x1	−0,074 38	0,997 23
x2	0,997 23	0,074 38

The eigenvalues are both negative, so stress is a maximum at the critical value. The contours of the predicted stress are ellipses centred at the critical value and with major axes in the directions given by the eigenvectors. The eigenvectors are almost proportional to (0,1) and (1,0), so the major axes are almost pointing in the directions of the coordinate axes. Three contour curves for the predicted stress are shown in Figure C.4. The information in the canonical analysis is that those contours are sections of ellipses centred right outside the lower left hand corner of the data range, more precisely at Spacing = 179,602 2 and Pressure = 7,891 2.

C.9.3 Application of the desirability function

The regression models for the three responses will be used to optimize the three responses simultaneously using two tools that are available in JMP®, the *prediction profiler* and the *contour profiler*. The prediction profiler will be illustrated in this Clause and the contour profiler will be illustrated in C.9.4.

The three responses are shown in the upper three rows of Figure C.3 and the response as a function of the two predictor variables are shown in the first two columns. The figure in the upper left corner shows the predicted Deposition Rate as a function of Spacing if Pressure is fixed at 8,5. The dashed lines above and below the predicted response are the 95 % confidence limits for the predicted response.

The prediction profiler can be used to obtain the optimal processing conditions via the desirability function. The right hand column of Figure C.3 shows how the experimental objectives are represented graphically as desirability functions.

— To maximize Deposition Rate, the desirability function in the upper right corner is 0 (undesirable) for values of Deposition Rate below 127 (the experimental objective), and 1 (most desirable) for high values of Deposition Rate.

— To minimize Non-Uniformity, the desirability function is 0 for high values of Non-Uniformity and 1 for low values (line going down).

— To take Stress to a target value of 1,5, the desirability function is 1 when Stress is 1,5, and decreases to 0 when Stress is higher or lower.

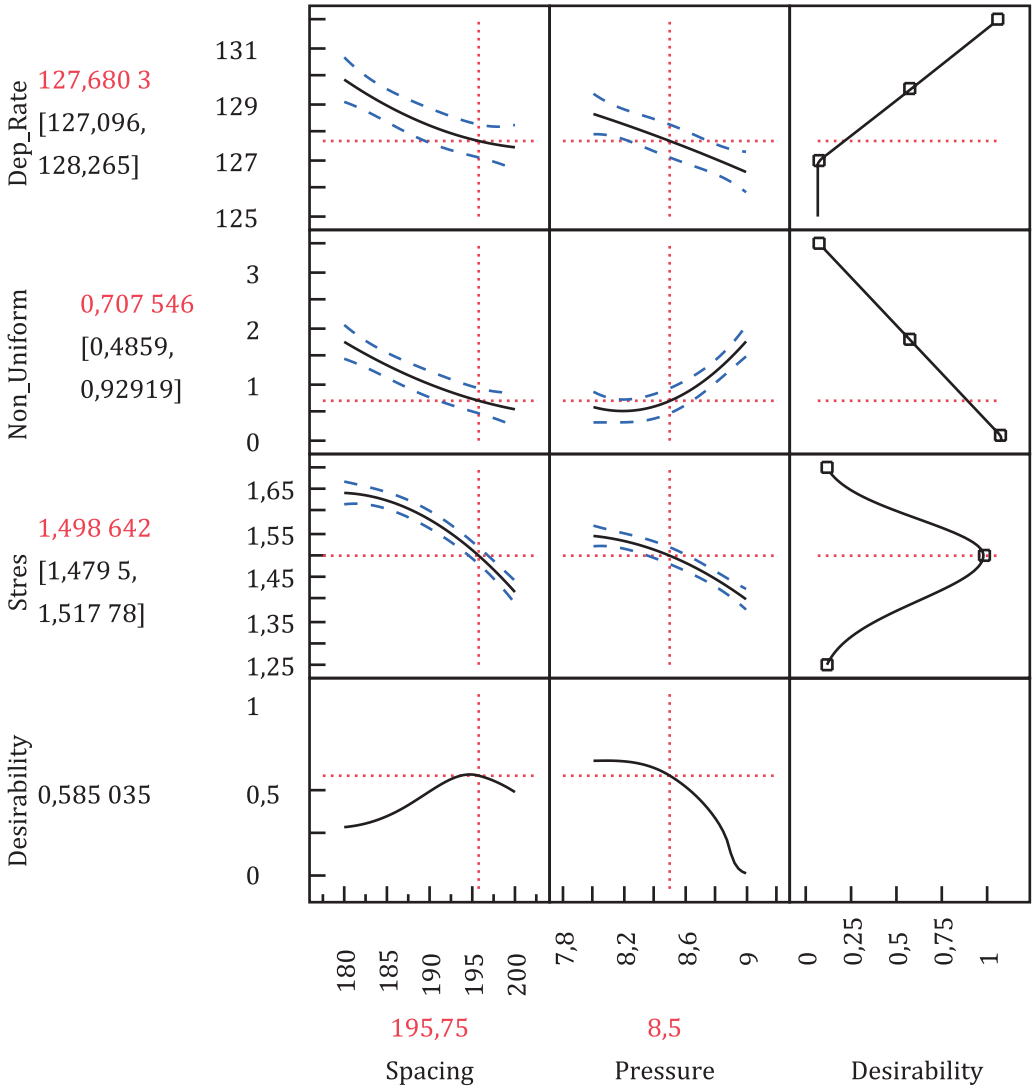


Figure C.3 — Prediction Profiler where the desirability function is used for optimization

With equal weight (importance) given to all three responses, the graph shows that the optimal factor settings are predicted to be

— Spacing = 195,75

— Pressure = 8,5

with response values of

— Deposition Rate = 127,7

— Non-Uniformity = 0,71

— Stress = 1,50

Note that this is a numerical optimization using the regression models for the three responses.

C.9.4 Application of contour plots to find optimal settings: Contour profiler

The contour profiler shown in Figure C.4 provides a better understanding of process behaviour and greater control over the factor settings during optimization.

The three cubes on the right show the three corresponding response surfaces as a function of the two factors in three dimensions. The graph at the bottom left shows overlaid contours of the three responses as a function of the predictor variables. A few details of the contours are as follows.

— The blue contour line for Stress is at the target value of 1,5. Two shaded areas were added with a Low Limit of 1,4 (shaded area in the lower left corner) and a High Limit of 1,6 (lower shaded area in the upper right corner), to indicate that the white area between them is close to target.

— The red area in the top right corner has a Deposition Rate below 127 (Low Limit), so the region below it satisfies the experimental objective of a Deposition Rate above 127. The red contour line going down from the upper left corner to the lower right corner is the 128,5 contour.

— The green contour lines for Non-Uniformity show that Non-Uniformity decreases in steps of 0,3 from 3,1 in the upper left corner (low Spacing and high Pressure) to 0,7 in the lower right corner. The green contour curve in the lower right corner without a label is the 0,49 contour.

The optimal conditions can be selected taking into account the relative importance of the three responses. Deposition Rate (speed at which oxide is deposited on the wafer) affects throughput, and is satisfactory above 127 Å/min. On the other hand, both Stress (with a target value), and Non-Uniformity (which reflects differences in oxide thickness across the wafer), affect quality. There needs to be a trade-off between Stress and Non-Uniformity. Two operating conditions will be considered:

— the optimal conditions where Stress is on target even though Non-Uniformity is not minimized, and

— the optimal conditions where Non-Uniformity is minimized even if Stress is slightly off target.

The optimal processing conditions where Stress is on target and Non-Uniformity is low are at the crosshairs where the contour line for Non-Uniformity touches the contour line for Stress, namely

— Spacing = 197,5;

— Pressure = 8,19;

where the predicted responses are as follows:

— Deposition Rate = 128,2;

— Non-Uniformity = 0,49;

— Stress = 1,50.

The optimal processing conditions that minimize Non-Uniformity within the region even though Stress is slightly off-target are

— Spacing = 200;

— Pressure = 8,25;

where the predicted responses are as follows:

— Deposition Rate = 127,9;

— Non-Uniformity = 0,45;

— Stress = 1,45.

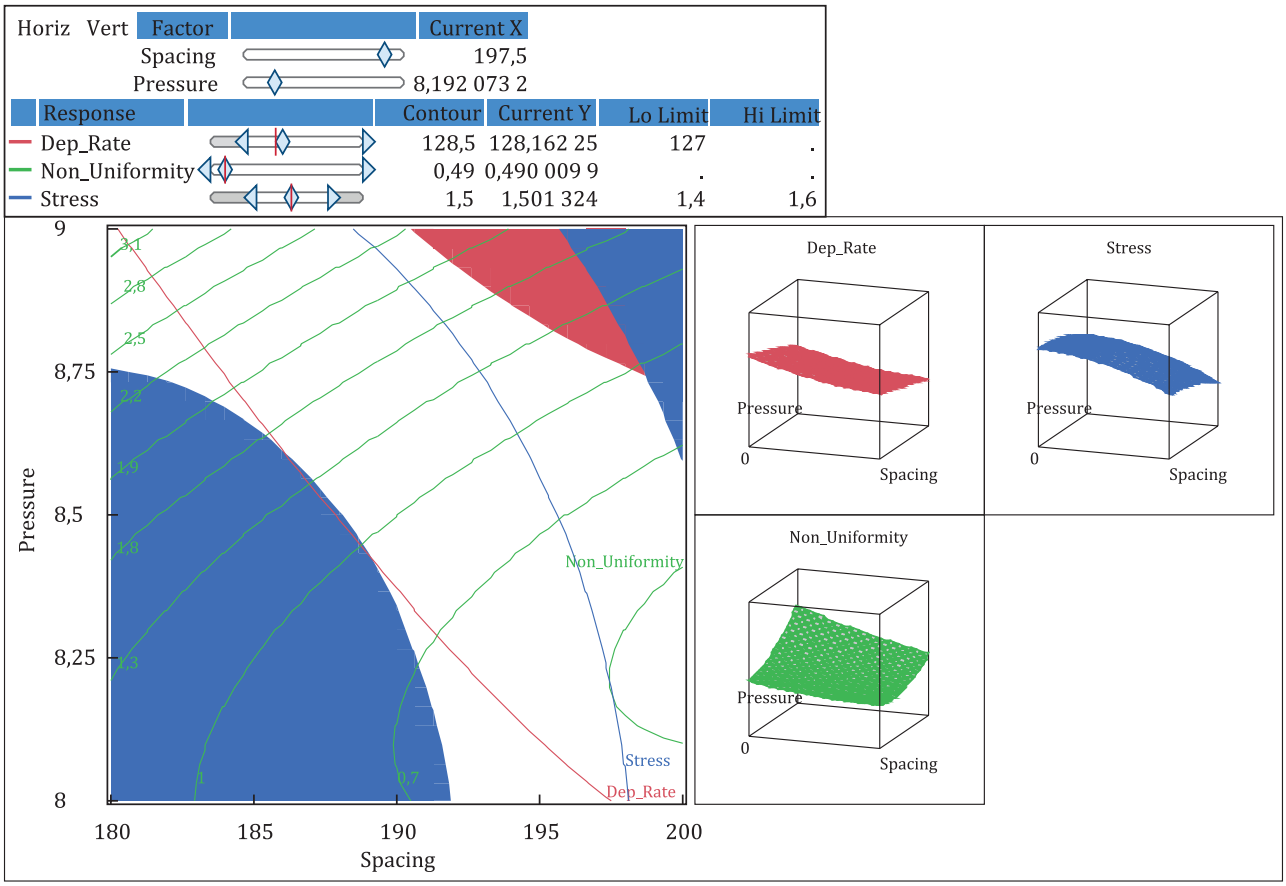


Figure C.4 — Contour profiler

C.9.5 Conclusion

The optimal factor settings for the PETEOS deposition process are the following:

— Spacing = 200 mils

— Pressure = 8,2 torr

at the following conditions.

— Non-Uniformity is predicted to be low (0,45 %) and robust, or insensitive, to changes in Pressure and Spacing.

— Stress is predicted to be 1,48 × 10^9 dynes/cm^2, which is close to the target value of 1,5 × 10^9.

— Deposition Rate is acceptably high at 128 Å/min.

The predictions at the optimal factor settings were confirmed using several reactor runs. To ensure process behaviour in the region around the new processing conditions was satisfactory, a new designed experiment, centred on the new processing conditions, was performed. The new processing conditions were adopted.

Annex D
(informative)

Process yield-optimization of a palladium-copper catalysed C-C-bond formation[5)]

D.1 Purpose of the experiment

D.1.1 General purpose

This example shows the yield optimization of the main product in a chemical reaction. The considered reaction is a so called Sonogashira-reaction that is applied to form carbon-carbon bonds in the synthesis of complex molecules, e.g. with endiyne units. Since those compounds can be cytotoxic by creating single and double stranded DNA cuts, the substance is needed in several fields of pharmacy and medicine. The purpose of the described experiment is to find optimum reaction conditions in preparation of a higher scale production of the product described in D.1.2.

The principles of the reaction are well known and understood. However, when applying specific substrates, such as the chosen base and solvent, the reaction needs a fine tuning to maximize the yield. The Sonogashira-reaction is a cross-coupling reaction between terminal acetylene and a vinyl halide under mediation of a Palladium (0)/Copper (I) catalyst. For the employment of the catalyst, it is expected that an increase of catalyst by a small amount will lead to an increase of the yield up to a specific optimum value. A further increase of catalyst is expected to decrease the yield. The same applies for the reaction time and temperature. Optimal values for these influential factors are known for similar reactions.

The assumed nonlinear behaviour encourages applying a CCD in order to locate the setting for an optimum yield based on a response surface.

D.1.2 Chemical background and experimental setup

The reaction formula is given below.

$$\underset{1}{X-R_1-X} + 2 \quad \underset{2}{H-\equiv -R_2} \Rightarrow \quad \underset{3}{R_2-\equiv -R_1-\equiv -R_2} \tag{D.1}$$

The used substrate (1) is 6,7-Dibromo-9,9-dimethyl-dioxa-spiro[4],[5]dec-6-ene.

The used substrate (2) is Trimethylsilylacetylene:

$$H-\equiv -Si(CH_3)_3 \qquad (2)$$

The main-product (3) is 9,9-Dimethyl-6,7-bis-silanyl-1,4-dioxa-spiro[4],[5]dec-6-ene.

This C(sp^2)-C(sp) bond formation process is catalysed by Pd(0) species, mostly $Pd(P(Ph)_3)_4$ and co-catalysed by Copper(I)iodine (CuI) in the anhydrous solvent benzene. Additionally, the reaction solution requires the presence of an amine base, and here piperidine was used. The reaction mixture comprises the following six components:

1. 6,7-Dibromo-9,9-dimethyl-dioxa-spiro[4,5]dec-6-ene [The substrate (1)];
2. Trimethylsilylacetylene [The substrate (2)];

5) Source: Thomas Pfeilsticker.

3. benzene as a solvent;
4. piperidine as a base;
5. Pd(0)tetrakis-(triphenylphosphin);
6. copper(I) iodine.

The experimental procedure is as follows.

Under inert atmosphere a 100 ml three-necked flask is charged with a solution of (1; (80 mmol)) in anhydrous benzene (freshly distilled), thermometer and refluxcondenser. Using inert gas technique the catalyst Pd(0)tetrakis-(triphenylphosphin) and the co-catalyst copper(I) iodine is added. After stirring the mixture for 30 min, anhydrous (distilled over CaH_2) piperidine is added within 10 min. Now 170 mmol Trimethylsilylacetylene (2), dissolved in anhydrous benzene, are added dropwise via septum technique to the stirred solution of (1).

At the end of the reaction, the reaction-mixture is cooled and diluted with saturated chlorine, and extracted with ether four times. The combined extracts are dried and concentrated under reduced pressure. The crude product is purified by column chromatography (SiO_2).

D.2 Response variable

Only one response variable was considered in this experiment: The process yield. The product is obtained as a colour-less oil, and the process yield is determined as a percentage (%) of the applied mass of substrates, see D.1.2.

D.3 Predictor variables

The Sonogashira-reaction which is being studied here is well known, so the experimenters were able to choose three predictor variables which were expected to have the strongest influence on the yield, and they were able to decide on feasible settings of those variables. The three variables were the following:

— **Reaction-time (R):** time inside the reaction flask in hours;

— **Temperature (T):** temperature during the reaction time in °C;

— **Concentration (C):** concentration of Palladium catalyst/copper(I) catalyst in mmol.

Although other predictor variables such as the amounts of solvent and base and the stirring speed were also expected to be active, it was decided to run an experiment where only the three predictor variables above were varied.

The chosen predictor variables can be varied in considerable ranges and a measurable result achieved.

The maximum value of the Reaction-time is 8 hours because the reaction has to be finished within one shift of the laboratory staff. In order to achieve a measurable amount of the main product, a minimum of 30 minutes is needed.

The maximum temperature is limited by the substrates. To avoid undesired side reactions, the temperature should not exceed 80 °C. A reasonable lower boundary is room temperature of 20 °C.

The minimum amount of Palladium catalyst is 5 mmol. In order to keep the cost for the reaction reasonable, 70 mmol (referred to Pd) should not be exceeded.

The description of the chemical reaction in D.1.2 mentions the co-catalyst Copper(I)iodine [Cu(I)], but the optimal Pd(0)/Cu(I) relation is well known in the literature and should be constant, thus Cu(I)-co-catalyst was set to 20,2 mmol Cu(I)iodine in the case of 10 mmol Pd(0), and to 50,4 mmol Cu(I)iodine in the case of 25 mmol Pd(0), for example.

D.4 Identification and estimation of measurement systems

The experiment will be conducted on laboratory scale. Thus, the measurement systems for controlling the predictor variables work under typical laboratory conditions. The systems are for the following:

— reaction time, R: simply an alarm clock, reminding the experimenter to stop the reaction;

— temperature, T: a thermometer with 0,01 °C resolution and 0,2 °C estimated expanded uncertainty of the measurement process;

— amount of catalyst, C: a weighing process resulting in a tested R&R value for mmol of 0,002 in the case of Pd(0).

These values have an influence on how precisely the settings of the predictor variables in the design can be realized. In order to assess whether the measurement systems are suitable, a reference is needed. In this example the defined factorial levels were used, see D.5.2. Since the reaction can be stopped in a few minutes there will be no problem with setting the values for R. Following the %R&R construction, relating 6·0,2 °C to the range of 25 °C for T (see D.5.2), the result will be 4,8 % which can be considered as acceptable. The same applies for the weighing process of Pd(0).

The process for the estimation of the yield comprises many steps and is more complicated. As stated in D.1.2, the reaction product is purified in a chromatography column. This separates the desired substance from all others by adsorption/desorption. Several compounds in the solution will be separated and appear in bands moving towards the bottom of the column. When the band of the desired product flows out, it needs to be detected and collected. Further steps for purification are to be carried out to isolate the desired main product from the solution. The mass of purified product is measured and set in relation to the amount of the applied reacting substrates. The measurement uncertainty for the combination of the weighing and purification processes is estimated to be ±7 % of the yield. With expected 80 % yield this is approximately ±5 to 6 % yield and considered to be quite high.

D.5 Selection of the settings of the predictor variables in the design

D.5.1 General considerations

A rotatable central composite design was chosen. Although for each factor variable a maximum range of variation is given in D.3 meaningful levels for factorial runs in the CCD are to be derived. Then the actual levels corresponding to the coded levels −1, 0, +1 are given. With the choice of a rotatable design $\alpha = 8^{1/4} = 1{,}682$ where 8 is the number of factorial runs. This means that $\alpha = 1{,}682$ is the distance in coded units from the middle level to the highest level, so the axial levels are found from the factorial levels as

$$l_1 = l_3 - 1{,}682 \cdot (l_3 - l_2) \text{ and } l_5 = l_3 + 1{,}682 \cdot (l_4 - l_3) \text{ with } C = (l_3 - l_2) = (l_4 - l_3)$$

This has to be taken into account when the levels l_2 and l_4 are to be defined. Since this is one of the most difficult decisions, the way how the values are derived here will be explained as follows.

From the most similar reaction the optimum point was used as a point of reference. This was 5 h for the reaction time, 35°C for the reaction temperature and 15 mmol for the amount of catalyst. Around this point the experimental region will be defined.

The optimum for the Reaction Time (T) is expected to be found within ±2 hours. That defines the centre level as l_3 = 5, and the factorial levels as l_2 = 3 and l_4 = 7. The axial levels are then calculated as Formula (D.2)

$$l_1 = 5 - 1,682 \cdot (5 - 3) = 1,636 \text{ and } l_5 = 5 + 1,682 \cdot (7 - 3) = 8,364. \quad \text{(D.2)}$$

However, the upper axial level of 8,364 hours exceeds the upper limit of 8 hours given in D.3. Therefore one hour will be subtracted from the levels found above. The new factorial levels are l_2 = 2 and l_4 = 6. The new axial levels are then derived as:

$$l_1 = 4 - 1,682 \cdot (4 - 2) = 0,636 \text{ and } l_5 = 4 + 1,682 \cdot (6 - 2) = 7,364 \quad \text{(D.3)}$$

The same approach was chosen for all other predictor variables.

The optimum temperature naturally depends on the substrates. Starting with 35 °C the best temperature for this reaction should be found within a range of approximately ±15 °C around this value. But in case of steric hindrance of the substance (1) higher reaction-temperatures may be useful. Thus, the centre of the experimental region is shifted by 10°C towards higher values. So the upper factorial level, l_4, is set to 60 °C. In order to keep the reference point within the experimental region, the lower factorial level ,l_2, is set to 35°C. The resulting levels are summarized in Table D1. It can be seen, that also the axial levels stay within the boundaries of operability given in D.4.

For the amount of catalyst l_3 is set to the reference value of 15 mmol. Within a range ±10 mmol around l_3 the optimum for this reaction is expected. By checking the axis levels again, it can be found that $l_1 = 15 - 1,682 \cdot (15 - 5) = -1,82$ mmol which cannot be realized. In order be able to achieve a measurable result from the experiment, l_2 was set to 10 mmol, and l_4 was kept at 25 mmol. All resulting levels are summarized in Table D1.

Special attention was given to ensure that in particular all combinations of extreme settings of the variables were within the region of operability.

All variables were continuous in the sense that all values within the chosen ranges were possible settings of the variables.

Table D.1 — Coded and actual levels of the experimental factors rounded to two digits after the decimal comma

Coded levels	−1,68	−1,00	0,00	+1,00	+1,68
Reaction time (R)	0,64	2	4	6	7,36
Temperature (T)	26,48	35	47,5	60	68,52
Concentration (C)	4,89	10	17,5	25	30,11

NOTE Coded levels are in the top row and actual levels are in the following rows. The units of the actual levels are hours for Reaction Time, °C for temperature and mmol for Palladium concentration.

Coded values of R, T and C are denoted by x_1, x_2, and x_3, respectively, and the conversion from the actual levels given in Table D.1 to the coded levels are given as

$$x_1 = \frac{\mathrm{R} - 4}{2} \quad \text{(D.4)}$$

$$x_2 = \frac{\mathrm{T} - 47,5}{12,5} \quad \text{(D.5)}$$

$$x_3 = \frac{C - 17,5}{7,5} \quad \text{(D.6)}$$

In a CCD, the settings or levels of the predictor variables are of three types to be considered next.

D.5.2 Factorial levels

The factorial levels are coded as –1 and +1, so the factorial levels in actual or natural units are given in the two columns labelled –1,00 and +1,00 in Table D.1. Factorial levels of the predictor variables in this experiment are as follows:

— **Reaction-time (R):** 2 and 6 (coded as –1 and +1 respectively);

— **Temperature (T):** 35 °C and 60 °C (coded as –1 and +1 respectively);

— **Concentration (C):** 10 mmol and 25 mmol (coded as –1 and +1 respectively).

The factorial levels have been given their name because those levels are used to define a factorial design, which is an important building block in a CCD. The factorial design is typically a full two-level factorial design or a fractional two-level design.

D.5.3 Axial levels

These axial levels are the extreme values the predictor variables can have, so the factorial levels in actual or natural units are given in the two columns labelled –1,68 and +1,68 in Table D.1. In this experiment, the axial levels are as follows:

— **Reaction-time (R):** 0,64 h and 7,36 h (coded as –1 and +1 respectively);

— **Temperature (T):** 26,48 °C and 68,52 °C (coded as –1 and +1 respectively);

— **Concentration (C):** 4,89 mmol and 30,11 mmol (coded as –1 and +1 respectively).

D.5.4 Centre levels

The centre levels are the average of the two factorial levels, and it is also the average of the two axial levels. Its coded value is 0. It was explained in D.5.1 how the centre levels were chosen with a view to keeping all five levels inside the maximal range of the predictor variables given in D.3. In this experiment, the centre levels are as follows:

— **Reaction-time (R):** 4 h (coded 0);

— **Temperature (T):** 47,5 °C (coded 0);

— **Concentration (C):** 17,5 mmol (coded as 0).

D.6 Experimental design

D.6.1 General

The levels of the individual predictor variables defined in D.5 are combined to define the runs of the experiment. The three types of runs in the CCD are the factorial, the star (or axial) and the centre runs.

Under the given laboratory conditions a high variation in the yield of the reactions is expected. Furthermore, as explained in D.4, the measurement process of the yield is expected to increase the variation. In order to increase the precision of the predictions from the experiment an unusual high number of runs for the central composite design were chosen and the factorial and axial runs are replicated twice.

D.6.2 Factorial runs

In the factorial runs, all predictor variables are at their factorial levels. The factorial runs usually constitute a full factorial or a fractional factorial design. In this experiment, a *replicated* full factorial design was chosen, so the number of factorial runs was $2 \times 2^3 = 16$ The factorial runs are given as serial numbers 1 to 16 in Table D.2. The factorial runs are given in actual levels in columns R, T, and C and in coded levels in columns x_1, x_2 and x_3.

D.6.3 Axial runs

In the axial runs, one of the predictor variables is at its upper or lower extreme level and the rest of the predictor values are at their centre value. The minimum number of axial runs with 3 predictor variables is 6, but in this design the experimenters decided to replicate the axial runs, so the total number of axial runs in the experiment is 12. The axial runs are given as serial number 25 to 36 in Table D.2.

D.6.4 Centre runs

In the centre runs all the predictor variables are at their centre level. Eight centre runs were used in this experiment. The eight centre runs are given as serial number 17 to 24 in Table D.2. The number of centre runs in this experiment is rather large. Both factorial and axial runs are replicated, so this large number of centre runs is not needed to provide an estimate of pure error.

Table D.2 — Experimental layout in actual levels of variables R, T and C and in coded variables x_1, x_2, and x_3.The yield in each run is recorded in the last column.

Serial order	Run order	R	T	C	x_1 (coded R)	x_2 (coded T)	x_3 (coded C)	Yield
1	28	2,00	35,00	10,00	–1,00	–1,00	–1,00	70,5
2	29	2,00	35,00	10,00	–1,00	–1,00	–1,00	69,8
3	6	2,00	35,00	25,00	–1,00	–1,00	1,00	68,3
4	18	2,00	35,00	25,00	–1,00	–1,00	1,00	67,3
5	15	2,00	60,00	10,00	–1,00	1,00	–1,00	70,4
6	31	2,00	60,00	10,00	–1,00	1,00	–1,00	68,9
7	1	2,00	60,00	25,00	–1,00	1,00	1,00	69,1
8	7	2,00	60,00	25,00	–1,00	1,00	1,00	68,1
9	23	6,00	35,00	10,00	1,00	–1,00	–1,00	74,2
10	33	6,00	35,00	10,00	1,00	–1,00	–1,00	74,5
11	21	6,00	35,00	25,00	1,00	–1,00	1,00	73,9
12	35	6,00	35,00	25,00	1,00	–1,00	1,00	74,1
13	8	6,00	60,00	10,00	1,00	1,00	–1,00	80,6
14	10	6,00	60,00	10,00	1,00	1,00	–1,00	81,2
15	5	6,00	60,00	25,00	1,00	1,00	1,00	80,2
16	32	6,00	60,00	25,00	1,00	1,00	1,00	77,9
17	4	4,00	47,50	17,50	0,00	0,00	0,00	77,5
18	12	4,00	47,50	17,50	0,00	0,00	0,00	75,2
19	14	4,00	47,50	17,50	0,00	0,00	0,00	77,6
20	20	4,00	47,50	17,50	0,00	0,00	0,00	76,5
21	22	4,00	47,50	17,50	0,00	0,00	0,00	75,6
22	27	4,00	47,50	17,50	0,00	0,00	0,00	76,6
23	30	4,00	47,50	17,50	0,00	0,00	0,00	74,4

Table D.2 *(continued)*

Serial order	Run order	R	T	C	x_1 (coded R)	x_2 (coded T)	x_3 (coded C)	Yield
24	34	4,00	47,50	17,50	0,00	0,00	0,00	79,2
25	3	0,64	47,50	17,50	–1,68	0,00	0,00	70,1
26	25	0,64	47,50	17,50	–1,68	0,00	0,00	69,1
27	2	4,00	26,48	17,50	0,00	–1,68	0,00	68,6
28	24	4,00	26,48	17,50	0,00	–1,68	0,00	70,0
29	9	4,00	47,50	4,89	0,00	0,00	–1,68	74,4
30	26	4,00	47,50	4,89	0,00	0,00	–1,68	75,7
31	11	4,00	47,50	30,11	0,00	0,00	1,68	72,6
32	36	4,00	47,50	30,11	0,00	0,00	1,68	72,3
33	16	4,00	68,52	17,50	0,00	1,68	0,00	75,5
34	19	4,00	68,52	17,50	0,00	1,68	0,00	74,7
35	13	7,36	47,50	17,50	1,68	0,00	0,00	79,9
36	17	7,36	47,50	17,50	1,68	0,00	0,00	80,3

D.7 Data generated by the experiment

The response variable from the experiment is the process yield and it is given in the last column in Table D.2. The product is obtained as a colour-less oil, and the process yield is determined as a percentage (%). The purpose of the experiment is to maximize yield. The actual levels of variables R, T and C and the coded variables x_1, x_2 and x_3 are given in the rows of Table D.2.

The second column of Table D.2 is the variable *run order* which gives the order in which the experimental runs were performed. The very important information from that column is first of all that the experiment was randomized. But it can also be seen that the replicates in the experiment are *genuine.* All the factorial and axial runs are replicated twice but only in one case are identical runs performed in sequence.

D.8 Analysis of results

D.8.1 Software used for the analysis

Data acquired through the experiment are analysed using destra®V11.

D.8.2 Transfer function

The estimated transfer function (the estimated systematic part of the model) for Yield (in %) is given by the following formula using actual or natural units

$$\begin{aligned}\text{Yield} = {} & 44{,}130 + 0{,}341 \cdot \text{R} + 0{,}893 \cdot \text{T} + 0{,}545 \cdot \text{C} - 0{,}172 \cdot (\text{R} \cdot \text{R}) + 0{,}057 \cdot (\text{R} \cdot \text{T}) - \\ & 0{,}010 \cdot (\text{T} \cdot \text{T}) + 0{,}010 \cdot (\text{R} \cdot \text{C}) - 0{,}0003 \cdot (\text{T} \cdot \text{C}) - 0{,}19 \cdot (\text{C} \cdot \text{C})\end{aligned} \tag{D.7}$$

The estimated coefficients are found in Table D.3.

The values of the predictor variables in actual units are rather large and so the estimated coefficient may become small. This is one argument for using coded units. The transfer function in coded units is as follows:

$$\begin{aligned}\text{Yield} = \; & 76{,}59 + 3{,}645 \cdot x_1 + \; 1{,}586 \cdot x_2 - 0{,}731 \cdot x_3 - 0{,}687 \cdot (x_1 \cdot x_1) \; + \; 1{,}413 \cdot (x_1 \cdot x_2) - \\ & 1{,}626 \cdot (x_2 \cdot x_2) + 0{,}150 \cdot (x_1 \cdot x_3) - 0{,}025 \cdot (x_2 \cdot x_3) - 1{,}077 \cdot (x_3 \cdot x_3)\end{aligned} \tag{D.8}$$

The estimated coefficients are found in Table D.4.

The analysis is performed in the next section with coded variables using the information in Table D.4 and the results are subsequently interpreted in terms of the actual variables. In the calculations that are reported in the following sections the axial point values are rounded as given in Table D.2.The conclusion of the analysis is that a second order model fits the data and that quadratic terms are needed in the model.

D.8.3 Estimation of coefficients

The result of the estimation is displayed in Table D.3 using actual levels and in Table D.4 using coded levels. It is the regression coefficients given in Table D.3 that are used to give the first representation of the transfer function in Formula (D.7) in D.8.2. It is not recommended to inspect the individual regression coefficients and trying to interpret whether some of the coefficient might fail to be significantly different from 0. The point is that a second order surface is being fitted and the least squares estimate of the fitted surface has the coefficients in the "b_i" columns of Table D.3 or Table D.4 depending on the coding of the predictor variables used. The column "s_{ci}" of Table D.3 and the column "s_{bi}" of Table D.4 show the standard error of the coefficients. From columns "b_i [...]", "$|t_i|$" and "P" can be seen, which coefficients are significantly different from 0. The asterisks in the text column "$|t_i|$" summarize the degree of significance. One asterisk means that the coefficient is significant using a significance level of 0,05, two asterisks mean that the coefficient is significant using a significance level of 0,01, and three asterisks mean that the coefficient is significant using a significance level of 0,005. The graphical column "$|t_i|$" shows a bar graph of "$|t_i|$" with the 0,05, 0,01, and 0,005 quantiles given as red lines. Thus it displays the information of significance graphically. The column "P" gives the commonly used *p*-value for the test statistic t. In column "b_i[...]", the 95 % confidence interval of the coefficients are shown. If the interval does not include the value 0 the coefficient is significantly different from 0 with significance level 0,05.

Table D.3 — Estimated regressions coefficients and CCD ANOVA table using actual units

Char.No	Char.Descr.	x_i	b_i	b_i [...]	s_{ci}	$\lvert t_i \rvert$	$\lvert t_i \rvert$	P
	Yield	$f(x_1 ..x_{10})$						
		Const.	44,13	33,42...54,84	5,209	8,471***		<0,000 1
	Reactiontime	A	0,341	-1,256...1,937	0,777	0,438		0,665
	Temperature	B	0,893	0,578...1,208	0,153	5,826***		<0,000 1
	Catalyst-concentration	C	0,545	0,104...0,986	0,214	2,542*		0,017 3
	Reactiontime	A^2	-0,172	-0,287...-0,057	0,055 9	3,078**		0,004 87
		AB	0,056 5	0,033 4...0,079 6	0,011 2	5,031***		<0,000 1
		AC	0,010 0	-0,028 5...0,048 5	0,018 7	0,534		0,598
	Temperature	B^2	-0,010 4	-0,013 3...-0,007 5	0,001 43	7,273***		<0,000 1
		BC	-0,000 3	-0,00 4...0,005 9	0,002 99	0,089		0,930
	Catalyst-concentration	C^2	-0,019 1	-0,027 3...-0,011 0	0,003 97	4,819***		<0,000 1

Table D.4 — Estimated regressions coefficients and CCD ANOVA table using coded levels

Char.No.	Char.Descr.	x_i	b_i	b_i	s_{b_i}	t_i	$\lvert t_i \rvert$	P
		Intercept	76,59	75,78...77,41	0,396	193,203***		<0,000 1
	Reactiontime	A	3,645	3,203...4,087	0,215	16,957***		<0,000 1
	Temperature	B	1,586	1,144...2,027	0,215	7,379***		<0,000 1
	Catalyst-concentration	C	-0,730	-1,172...-0,289	0,215	-3,398**		0,002 19
	Reactiontime	A^2	-0,688	-1,148...-0,229	0,224	-3,078**		0,004 87
		AB	1,412	0,835...1,990	0,281	5,031***		<0,000 1
		AC	0,150	-0,427...0,727	0,281	0,534		0,598
	Temperature	B^2	-1,624	-2,083...-1,165	0,223	-7,273***		<0,000 1
		BC	-0,025 0	-0,602 1...0,552	0,281	-0,089		0,930
	Catalyst-concentration	C^2	-1,076	-1,535...-0,617	0,223	-4,819***		<0,000 1

It is worth pointing out that the estimates of the regression coefficients depend on the coding of predictor variables used. Note for example that the estimated coefficient of R is 0,341 and it is non-significant when actual levels are used (see Table D.3), while the estimated coefficient of x_1 is 3,645 and is strongly significant when coded levels are used.

While it is inappropriate to change the values of individual estimated regression coefficients it does make sense to look for model simplification where one answers questions like the following:

— Are the pure quadratic terms needed in the model?

— Is a first order model satisfactory?

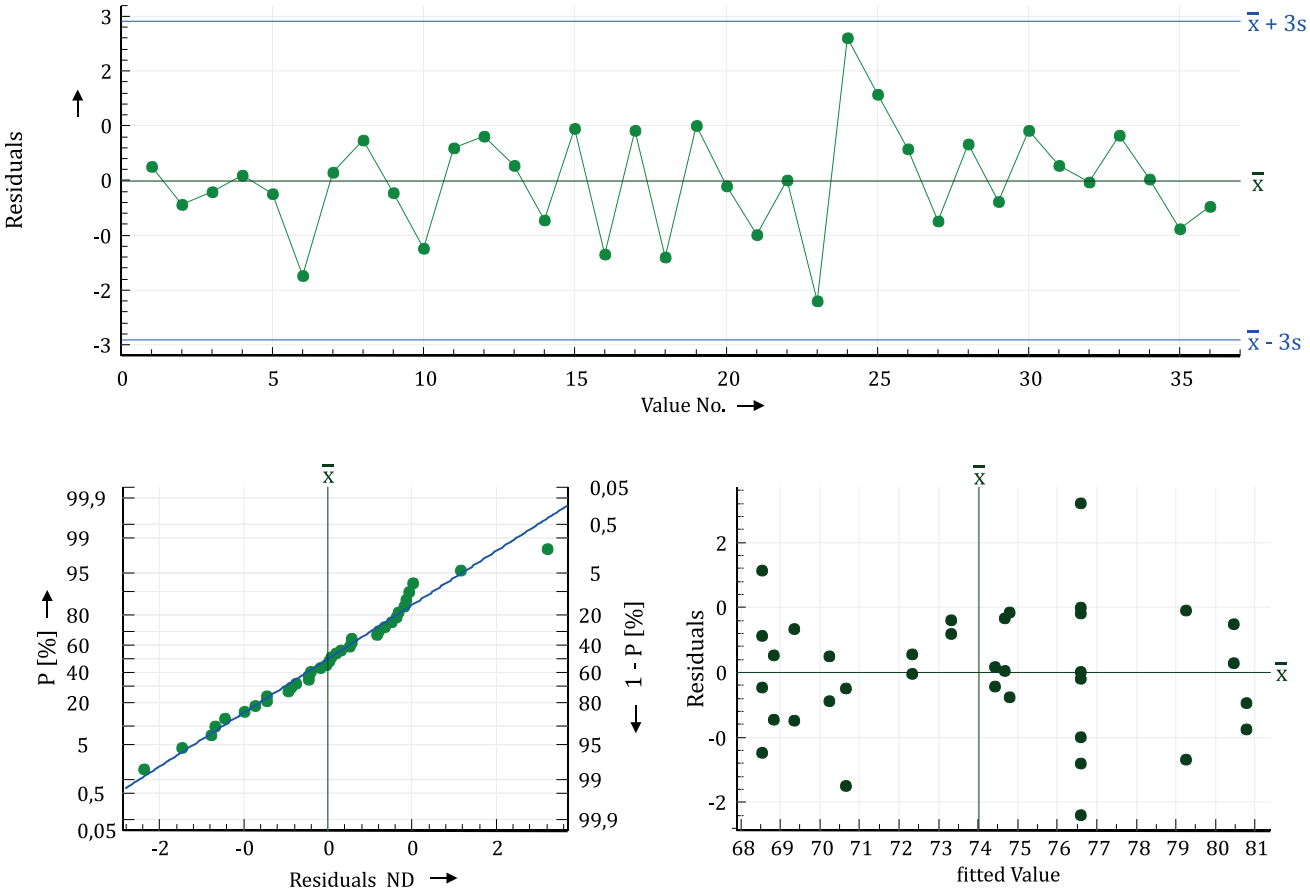
This is where the ANOVA table in Table D.5 is useful. The ANOVA tables are the same regardless of which coding of the variables is being used. The first 12 lines in the table give the information based on the F-test whether a specific parameter β_i, β_{ij}, β_{ii} or a combination of them is needed in the model. The line "Quadratic" indicates the need for pure quadratic terms in the model by giving the F-test for the hypothesis that $\beta_{11} = \beta_{22} = \beta_{33} = 0$ in the F_0-column. The p-value of the F-test is smaller than 0,0001 and the hypothesis is strongly rejected, so the quadratic terms are needed in the model. The F-test in the "Two-factor-interaction" line tests the hypothesis $\beta_{12} = \beta_{13} = \beta_{23} = 0$ and this hypothesis is also rejected and so is the hypothesis $\beta_1 = \beta_2 = \beta_3 = 0$ which is tested by the F-test in the "Linear" line.

Table D.5 — CCD ANOVA table using coded levels

R^2= 94,474 %						R^{*2}= 92,561 %		
Char.No.	Char.Descr.	x_i	SS	MS	DF	F_0	F_0	P
		linear	445,9	148,6	3,000	117,9***		<0,000 1
	Reactiontime	A	362,6	362,6	1,000	287,6***		<0,000 1
	Temperature	B	68,67	68,67	1,000	54,45***		<0,000 1
	Catalyst-concentration	C	14,56	14,56	1,000	11,55**		0,002 2
		quadratic	82,39	27,46	3,000	21,78***		<0,000 1
	Reactiontime	A^2	11,95	11,95	1,000	9,474**		0,004 9
	Temperature	B^2	66,72	66,72	1,000	52,90***		<0,000 1
	Catalyst-concentration	C^2	29,28	29,28	1,000	23,22***		<0,000 1
		Two-factor interaction	32,29	10,76	3,000	8,535***		0,000 4
		AB	31,92	31,92	1,000	25,31***		<0,000 1
		AC	0,360	0,360	1,000	0,285		0,598
		BC	0,010 0	0,010 0	1,000	0,007 93		0,930
		Model	560,6	62,28	9,000	49,39***		<0,000 1
		Error	32,79	1,261	26,00	---		---
		Lack of Fit	8,384	1,677	5,000	1,443		0,250
		Pure Error	24,40	1,162	21,00	---		---
		Total	593,3	16,95	35,00	---		---

The last five lines concern the error estimate and a model check. The experiment has 36 observations and the model has 10 parameters, one for each of the 10 terms on the right hand side of the formulae for the transfer function in D.8.2. This gives an error estimate with 36 – 10 = 26 degrees of freedom (see the line "Error"). Because of the replications, 21 degrees of freedom can be identified for pure error (see line "Pure Error"). The 21 degrees of freedom are 7 degrees of freedom from the 8 replicates at the centre point and 14 degrees of freedom from the two replicates at each of the 14 design points (8 factorial points and 6 axial points). Calculating the sum of squares for Pure Error (line "Pure Error", column "SS") and subtracting it from the Total Error Sum of Squares (line "Error", column "SS") gives the Lack of Fit Sum of Squares in the line "Lack of Fit". Dividing it by the remaining 5 degrees of freedom for the lack of fit, leads to the mean squares for Lack of Fit in column "MS". The "Lack of Fit" F-test is the ratio of the mean square for Lack of Fit to the mean square for Pure Error. In this case the *p*-value is 0,250 and the model is not questioned. The test statistic is given in column "F_0" together with the asterisks showing the degree of significance by comparing F_0 with the F-quantiles for the α-values as stated above. The graphical "F_0"-column shows again the comparison of F_0 with F-quantiles for the different α-values. The filled diamond on left side in graphic in line "Lack of Fit" indicates that F_0 falls below the F-quantile for α = 0,05. This leads to the same conclusion from the *p*-value: No significant Lack of Fit.

D.8.4 Graphs of data and residuals



NOTE The plot in the upper panel shows the residuals against the serial order from the first column of Table D.2. The type of run can be identified on the plot noting that runs 1 to 16 are factorial runs, runs 17 to 24 are centre runs, and runs 25 to 36 are star runs. The plot in the lower left hand panel shows a Q-Q plot of residuals. Finally the plot in the lower right hand panel shows the residuals plotted against the fitted value.

Figure D.1 — Selection of plots of the data and the residuals

The plot in the upper panel shows the residuals against the serial order of the runs from the first column of Table D.2. The plot gives an overview over the residuals and shows their ±3 s range. The eye-catching highest deviations from the model occur for No. 23 and 24 which are the last two replications of the centre points. The overall impression is that the deviations are equally spread around the model and no systematic trend is present which would indicate a systematic lack of fit. Comparing columns one and two of Table D.2 it is seen that the type of run can be identified from the serial order. Thus, the first 16 points are the pairs of replicated factorial runs, they are followed by 8 replicated centre runs, and, finally, by 12 replicated star runs.

The plot in the lower left panel shows a Q-Q plot of residuals as a check for normal distribution. This plot does not contradict the assumption of a normal distribution. The two largest residuals in runs 23 and 24 show up again. But they are not critical for the assumption of a normal distribution. Larger deviations from the straight line in a Q-Q plot are always expected for smaller and for larger observations.

In the plot in the lower right panel, the residuals are plotted against the fitted values. The values in the boxes show again the serial number in the design. If, for example, large standard deviations occurred with large responses, this plot would show a funnel shape. But again, only randomness can be seen. For the fitted value of 76,5 %, the replications in the centre point can easily been identified and the large residual in runs 23 and 24 are noted again.

Comparing the range of residuals of less than ±3 % yield with a range of the explained yield in the experimental region of about 13 % the plot also indicates a good fit of the model as already noted in the ANOVA tables in Table D.5.

D.9 Presentation of results — Optimization

D.9.1 General

Just looking at the transfer function (the estimated systematic part of the model) whether it is shown in actual units as in Formula (D.7) or in coded units as in Formula (D.8) does not reveal the properties of the response surface. It is important to realize that the second order polynomial in two or more variables can represent a variety of surfaces. The stationary point can be a maximum, a minimum or a saddle point. Furthermore, the stationary point can be located outside as well as inside the experimental region. The situation is only simple, if the objective is to maximize the response and if the stationary point is a maximum located inside the experimental region, or, alternatively, if the objective is to minimize the response and if the stationary point is a minimum located inside the experimental region, and then the optimal settings is the stationary point. In all other cases, the optimal setting needs to be found away from the stationary point but within the boundary of the experimental region. If the number of predictor variables is two or three, then contour plots can be of help as illustrated in D.9.3.

D.9.2 presents the canonical analysis of the fitted second order response surface. Canonical analysis has been the basic tool to gain a precise understanding of the response surface. Although modern computer software has various impressive graphical and analytical tools, the knowledge obtained from a canonical analysis helps to fully exploit the graphical tools. The results of the canonical analysis in this experiment are that the stationary point is a maximum that is located far outside the experimental region. The analysis is not done with a user function of destra® V11.

In addition to the canonical and the graphical analyses, many software packages easily apply numerical search methods in order to find a local minimum or maximum of the response surface. D.9.3 gives the result of the search by destra® V11 applying Powell's search algorithm within the experimental region.

D.9.2 Canonical analysis

D.9.2.1 Eigenvalues, eigenvectors and stationary point

The standard result from the canonical analysis based on any software package is given is Table D.6. Table D.7 gives the coordinates of the stationary point, $\boldsymbol{x}_S$, in coded units and in original units as well as the predicted response at the stationary point. Mathematical details of canonical analysis can be found in E.3.

Table D.6 — Canonical analysis of response surface based on coded variables: Stationary point

	Stationary point		
	R	**T**	**C**
Coded units	5,697	2,962	0,023
Original units	15,380	84,531	17,671
Predicted value at the stationary point			
89,30			

Table D.7 contains information that characterizes the estimated response surface. The first column gives the eigenvalues and all three are negative, which means that the stationary point is a maximum. The eigenvector corresponding to the eigenvalue −0,304, for example, is given in the same row. The labels R, T and C have been used for the coordinates even if the coordinates are in coded units. All eigenvalues are negative so the stationary point is a maximum.

Table D.7 — Canonical analysis of response surface based on coded variables: Eigenvalues and eigenvectors

Eigenvalues	Eigenvectors		
	R	T	C
$\hat{\lambda}_1 = -0{,}304$	0,880	0,469	0,078
$\hat{\lambda}_2 = -1{,}079$	−0,045	−0,081	0,996
$\hat{\lambda}_3 = -2{,}007$	−0,474	0,879	0,050
Stationary point is a maximum			

The distance of the stationary point to the design centre in coded units is

$$D_S = 6{,}42.$$

So the stationary point is, in this case, rather far outside the experimental region because the extreme points which are the factorial points and the axial points in this design are at a distance from the design centre of $\sqrt{3} = 1{,}73$ and $\alpha = 1{,}68$, respectively. The stationary point is a maximum so an optimum must be sought near the boundary of the experimental region and as close as possible to the stationary point.

The estimated response at the centre of the experimental region is 76,59 with a standard error of 0,396 as can be seen from Table D.4, because the estimated response at the design centre is the intercept when coded variables are used. The estimated maximal response is 89,30 with a standard error of 9,08. This rather large standard error underlines that extrapolation outside the experimental region is futile.

D.9.2.2 Finding a recommended setting

In this example it is fairly easy to find points in the experimental region with higher estimated response than at the design centre. The vector $\boldsymbol{x}_S = (5{,}697,\ 2{,}961,\ 0{,}023) = 6{,}4 \cdot (0{,}890, 0{,}463, 0{,}004)$ is approximately proportional to the eigenvector of the numerically smallest eigenvalue $(0{,}880, 0{,}469,\ 0{,}078)$, see Table D.7. This means that moving away from the centre point in increments proportional to $\boldsymbol{x}_S$ will be the increase the response most quickly.

The point (1,54, 0,80, 0,006) has a distance of 1,74 from the design centre so it is located on the boundary of the experimental region. The expected response with these settings is 82,54 with a standard error of 0,652 41. In original units the recommended setting is (R,T,C) = (7,08 h, 57,5 °C, 17,55 mmol).

D.9.2.3 Is the stationary point of the underlying true surface also a maximum?

There is no doubt that the estimated response surface has a single global maximum, and there is no doubt that the fit of the second order model is satisfactory as explained in D.8.2 and reported in the ANOVA table in Table D.5. But the eigenvalues, $\hat{\lambda}_i$, in Table D.7 that determine the properties of the estimated response surface are only estimates of the eigenvalues, λ_i that determine the properties of the underlying true surface. The estimated eigenvalues in Table D.7 are functions of the estimated coefficients of the transfer function and they are, therefore, subject to error. Their standard errors are roughly of the same size as those of the quadratic coefficients of the transfer function (0,223), see Table D.4. The estimated eigenvalues and their standard errors can provide confidence intervals for the

eigenvalues of the underlying true surface by the usual formulas. The formula for the 95% confidence interval for λ_i is

$$\hat{\lambda}_i - se(\hat{\lambda}_i) \cdot t_{0,975}(v) \leq \lambda_i \leq \hat{\lambda}_i + se(\hat{\lambda}_i) \cdot t_{0,975}(v)$$

where v denotes the degrees of freedom for total error in the experiment. In this experiment $v = 26$ as can be seen in the first ANOVA table in Table D.5, so $t_{0,975}(v) = t_{0,975}(26) = 2,0555$. The 95% confidence intervals for the three eigenvalues are as follows:

$$\begin{aligned} -0,304 - 0,223 \cdot (2,0555) \leq \lambda_1 &\leq -0,304+0,223 \cdot (2,0555) \\ -0,762 \leq \lambda_1 &\leq 0,154 \end{aligned} \tag{D.9}$$

$$\begin{aligned} -1,709 - 0,223 \cdot (2,0555) \leq \lambda_2 &\leq -1,709+0,223 \cdot (2,0555) \\ -1,537 \leq \lambda_2 &\leq -0,621 \end{aligned} \tag{D.10}$$

$$\begin{aligned} -2,007 - 0,223 \cdot (2,0555) \leq \lambda_3 &\leq -0,2,007+0,223 \cdot (2,0555) \\ -2,466 \leq \lambda_3 &\leq -1,549 \end{aligned} \tag{D.11}$$

The 95% confidence interval for the numerically smallest eigenvalue, λ_1, contains 0 and extends into the positive numbers. While it is too pessimistic to conclude that the true surface has a saddle point rather than a global maximum, it does mean that there potentially is a slower increase in response as the settings are changed in the direction of the eigenvector corresponding to λ_1, and it is exactly in that direction that the recommended settings were found.

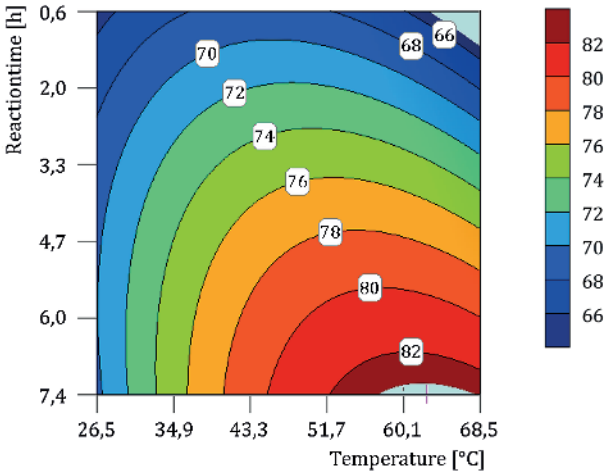
D.9.3 Application of contour plots to find optimal settings

With three predictor variables a series of contour plots of the response surface in two variables for a range of appropriately chosen values of the third value are needed to explore the response surface.

The contour plots in Figure D.2 are slices cutting through the contours in three-dimensional space at the selected values of C. The fixed values in the first three contour plots are selected as the centre point as well as the extreme values of C that are used in the experiment, see Table D.1. The centre point is shown in the contour plots as a white cross that can be moved by sliding the red vertical lines in the three small graphics on the right side of the plot.

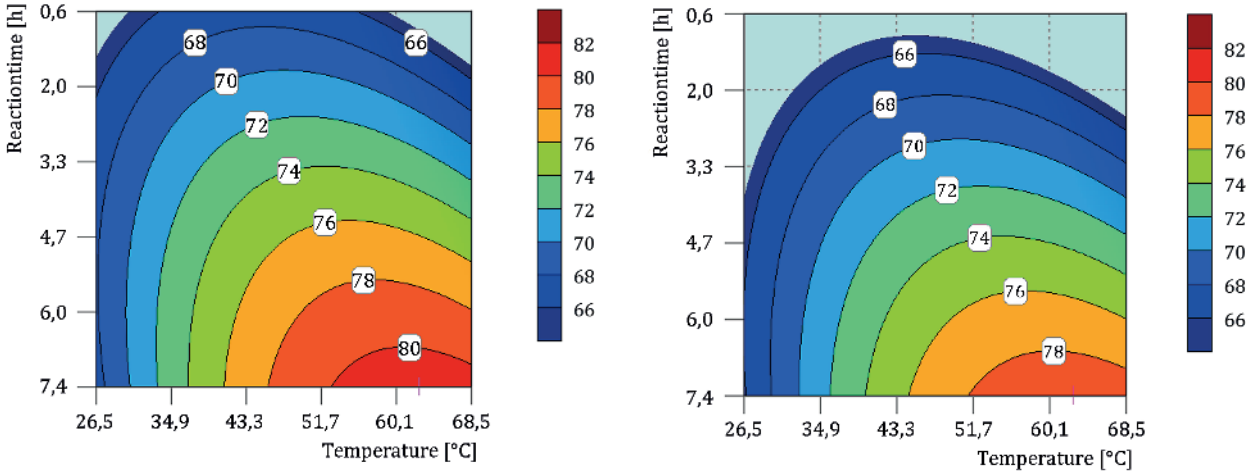
Just looking at the contour plots it is obvious that the high responses are found for large values of both R and T regardless of the value of C. The highest response seems to be found near a setting of C = 17,5 mmol which was the central value of C in the experiment. The experimenter who is basing the search for optimal settings on contour plots alone may feel a need for further contour plots with values of C around 17,5. This can easily been displayed by sliding the red line for variable C in the software graphic. The lower contour plots show the response surface for C = 5 mmol and C = 30mmol. It is obvious that the optimum for C should be in-between.

C = 17,5 mmol:



C = 5 mmol:

C = 30 mmol:



NOTE The top subplot shows the contour plot with C at its centre value C = 17,5. The lower two plots show the contour lines for C close to its extreme values.

Figure D.2 — Contour plots of predicted responses for three fixed values of C

In order to find the optimum graphically with the software, the experimenter can change C by setting the slider around the maximum of the red function (the cut through the response surface along C at values as set for R and T) in the respective right lower graphic for C.

Figure D.3 shows the location of the maximum of yield within the experimental region. The point can be found by R = 7,3 h, T = 63 °C and C = 16 mmol. The found optimum is obviously located at the border of the experimental region for the reaction time R. Thus, it can be concluded, that a longer reaction time, maybe exceeding 8 h, will lead to the maximum of yield.

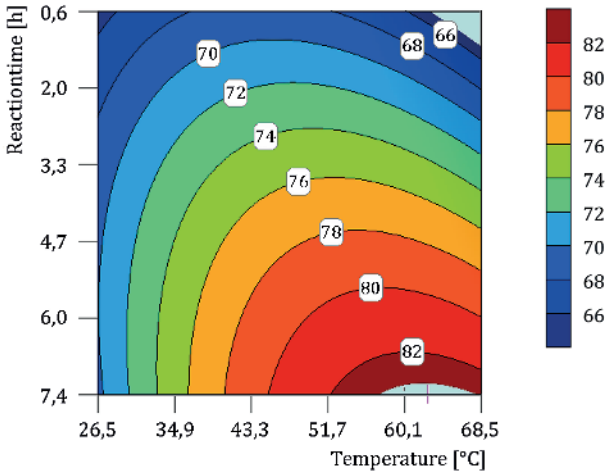


Figure D.3 — Contour plot showing the location of maximum within the experimental region marked with the red cross at R = 7,36 h, T = 62,76 °C, C = 15,73 mmol

D.9.4 Numerical search to find optimal settings

The search for the location of the optimum can of course been numerically done by the software. This seach is carried out within the experimental region. The maximum point is as follows:

— R = 7,36 h

— T = 62,76 °C

— C = 15,73 mmol

— Yield = 83,25 % ; Standard error for the yield: 1,45 %

This analysis is based on the identified response surface but only within the experimental region. The goodness-of-fit of the model has only been assessed for settings within the experimental region. If the optimum point is found outside this region, accepting this point as identified optimum presumes that the model approach is true. This is not necessarily the case in practice. Often the response surface gives only a good approximation within the investigated region.

It is natural to compare the recommended setting given in D.9.2.2 with the so-called optimal setting found by numerical search in this clause. The recommended setting in D.9.2.2 is (R,T,C) = (7,08 h, 57,5 °C, 17,55 mmol) with an expected yield of 82,54 % and a standard error of 0,65241 %. This setting has a slightly lower expected yield and a substantially lower standard error than the optimal setting of this clause. The explanation is that the optimal setting of this clause is further away from the design centre. In coded units, its distance from the design centre is 2,09 while the recommended setting of D.9.2.2 is 1,74 units from the design centre. It could be argued that the design region is a ball[6)] centred at the design centre and a radius equal to the larger of α and the distance of the factorial points from the design centre. In this case the radius is 1,74. With this definition of the design region, the so-called optimal point is located outside the experimental region.

6) There seems to be some confusion concerning the meanings of a **sphere** and a **ball**. In mathematics, a sphere is defined as the set of points that are all the same distance, r (the radius), from a given point (the centre) in three-dimensional space. A **ball** is the inside of a sphere. It may be a **closed ball** (including the sphere) or an **open ball** (excluding the sphere). These concepts are defined not only in three-dimensional space but also for lower and higher dimensions. In two-dimensional space, the plane, a ball is the same as a disc, and a sphere is the same as a circle. Informally, a **solid sphere** is used as a synonym for a ball.

D.9.5 Conclusion and confirmation run

The experiment has been undertaken on a laboratory scale in preparation of up-scaling the reaction for a mass production. As a result an optimum has been found under restriction of eight hours reaction time. For the used substrates in the reaction it has been shown where the optimum concentration of the palladium-copper catalyst is located and what the optimum temperature is.

These findings can be taken as starting point for settings in mass production. Usually, conditions under up-scaled mass production differ from those in the laboratory. Thus, a further adjustment of the settings will necessarily be done in the production processes applying other optimization techniques (e.g. EVOP[3],[4]). For this reason, no confirmation runs have been made.

Although no confirmation runs were made, the 95 % prediction interval that should have been used to evaluate the predictions of the model is still interesting. If the predictions of the experiment are to be confirmed, 95 % of the future runs should give yields inside the 95 % prediction interval. The formula for the 95 % prediction interval is

$$\hat{y} - t_{0,975}(v) \times \sqrt{\text{std.dev}^2 + \text{std.err}^2} \leq \textit{yield} \leq \hat{y} + t_{0,975}(v) \times \sqrt{\text{std.dev}^2 + \text{std.err}^2}$$

where $\hat{y}$ is the predicted value at the chosen settings, v are the degrees of freedom for error, std.dev is the estimated standard deviation of an observation and std.err is the estimated standard error of the $\hat{y}$.

For the optimal settings found by the numerical search in D.9.4 the values are $\hat{y} = 83,25$, v = 26, $\text{std.dev}^2 = 1,263$ (calculated as the Error SS divided by v), and std.dev = 1,45. The result is

$$83,25 - 2,06 \times \sqrt{1,263 + 1,45^2} \leq \textit{yield} \leq 83,25 + 2,06 \times \sqrt{1,263 + 1,45^2},$$
$$79,5 \leq \textit{yield} \leq 87,0$$

This interval is rather wide.

Annex E
(informative)

Background on response surface designs

E.1 Sequential assembly of central composite designs for two predictor variables

The second order model in two predictor variables x_1 and x_2 is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} {x_1}^2 + \beta_{22} {x_2}^2 + \varepsilon \quad \text{(E.1)}$$

A central composite design is shown in the Table E.1. The columns with headers x_1 and x_2 contain the setting of the two factors in coded units. The first column is just an indexing of the runs for easy reference.

Table E.1 — Central composite design in two predictor variables

Run id	x_1	x_2
1	–1	–1
2	1	–1
3	–1	1
4	1	1
5	0	0
6	0	0
7	–1,41	0
8	1,41	0
9	0	–1,41
10	0	1,41

The shading of the rows in Table E.1 illustrates the three different types of runs in the design. The first four runs with the lighter shading are factorial points, the following two runs are centre points, and the last four runs are star points.

The model is fitted using linear regression with a set of predictor variables given in Table E.2. In addition to the design variables from Table E.1, x_1 and x_2, the predictor variables are a column of ones, denoted by e, the product of x_1 and x_2, denoted by x_1x_2, and the squares of x_1 and x_2 denoted by x_1^2 and x_2^2. The matrix in Table E.2 is called the design matrix in regression terminology, but when applied to a designed experiment this is a misnomer; for the design is determined by x_1 and x_2 and the rest of the variables in Table E.2 are calculated from x_1 and x_2 and they are added because the model is Formula (E.1). The term model matrix is also used for the matrix in Table E.2 and it makes more sense.

The shading corresponding to the three types of runs in a central composite design are kept in Table E.2.

The shading is also related to the sequential assembly of the central composite designs. Quite often, an experimenter starts out with many factors and therefore first performs a screening experiment to investigate which factors are locally active. This screening experiment is often a fractional factorial

and because of the projectivity properties of fractional factorials the experimenter ends up with a full factorial in a few factors. This is illustrated here with the full factorial design in the two two-level factors in the first four runs.

The model that can be fitted with this design is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

Table E.2 — Columns 2 to 7 are the model matrix for the second order model based on the central composite design given in Table E.1

Run id	e	x_1	x_2	x_1x_2	x_1^2	x_2^2	Y
1	1	−1	−1	1	1	1	y_1
2	1	1	−1	1	1	1	y_2
3	1	−1	1	−1	1	1	y_3
4	1	1	1	−1	1	1	y_4
5	1	0	0	0	0	0	y_5
6	1	0	0	0	0	0	y_6
7	1	−1,41	0	0	1,988 1	0	y_7
8	1	1,41	0	0	1,988 1	0	y_8
9	1	0	−1,41	0	0	1,988 1	y_9
10		0	1,41	0	0	1,988 1	y_{10}

If the experimenter has the second order model in mind but only has data from the factorial points of the design, it is clear that only four parameters can be estimated and if four parameters of the mean are estimated there is no estimate of the error variance. A look at the first four rows of the matrix in Table E.2 corresponding to the factorial points reveals that the columns e, x_1, x_2, and x_1x_2 are different, but e, x_1^2 and x_2^2 are identical. This means that with the data from the factorial points only the parameters of the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \qquad \text{(E.2)}$$

can be estimated. But it also means that if estimation is performed in the first order model Formula (E.2) and the true model is the second order model Formula (E.1), then the mean of b_0, the estimate of β_0, is

$$\beta_0 + \beta_{11} + \beta_{22}$$

Now, if the experimenter adds two centre points, the design matrix will be the first 6 rows of Table E.2. The columns e, x_1, x_2, and x_1x_2 are still different, but now e is different from x_1^2 and x_2^2 which are still identical. This means that the experimenter from these 6 runs can obtain unbiased estimates of β_0, β_1, β_2, β_{12} and of $\beta_{11} + \beta_{22}$. This is important, for the extra centre runs make it possible for the experimenter to evaluate whether it is worthwhile to perform extra runs in order to fit the full second order model Formula (E.1).

If this is the, case the experimenter can add the last four runs of Table E.1 and now all the columns of the design matrix of Table E.2 are different and all the parameters of the second order model Formula (E.1) can be estimated.

E.2 The versatility of second order models

The success of the second order model in response surface methodology stems from the fact that very different surfaces can be approximated by second order polynomials. This can be illustrated with just two predictor variables with an example taken from Chapter 11.1 of Reference [2].

Choosing the coefficients in the second degree polynomial in the deterministic part of the second order model Formula (E.1) four polynomials representing estimated responses can be obtained as follows:

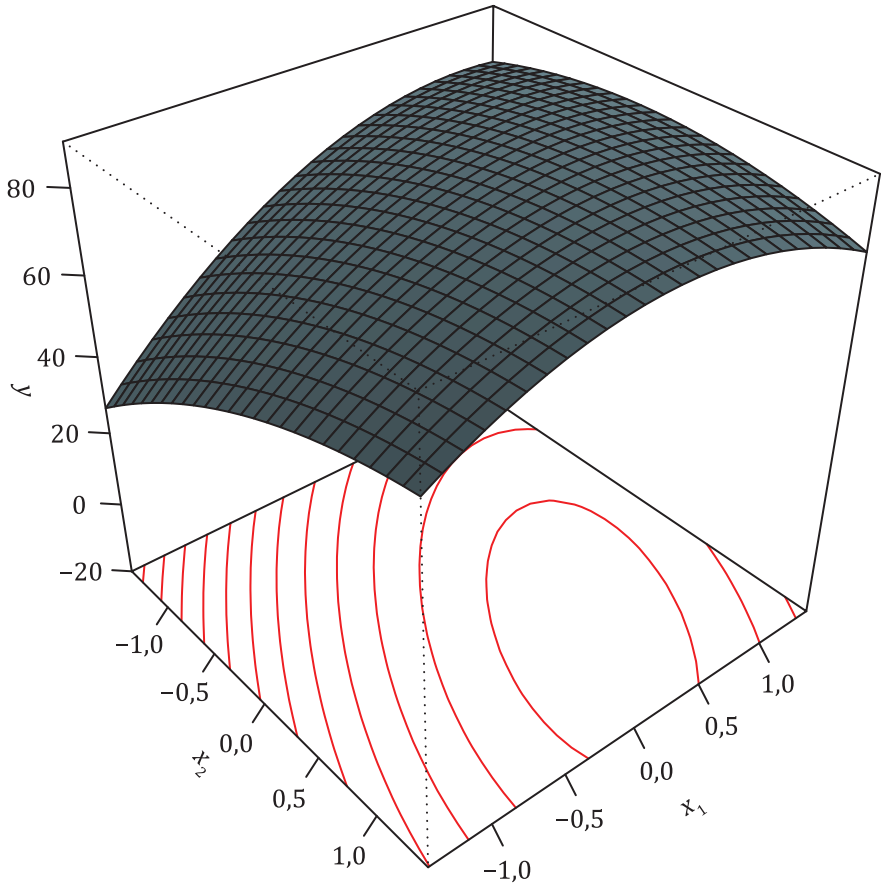
$$\hat{y} = 83{,}6 + 9{,}4x_1 + 7{,}1x_2 - 5{,}8x_1x_2 - 7{,}4x_1^2 - 3{,}7x_2^2 \quad \text{(E.3)}$$

$$\hat{y} = 83{,}9 + 10{,}2x_1 + 5{,}6x_2 - 7{,}6x_1x_2 - 6{,}9x_1^2 - 2{,}0x_2^2 \quad \text{(E.4)}$$

$$\hat{y} = 82{,}7 + 8{,}8x_1 + 8{,}2x_2 - 7{,}6x_1x_2 - 7{,}0x_1^2 - 2{,}4x_2^2 \quad \text{(E.5)}$$

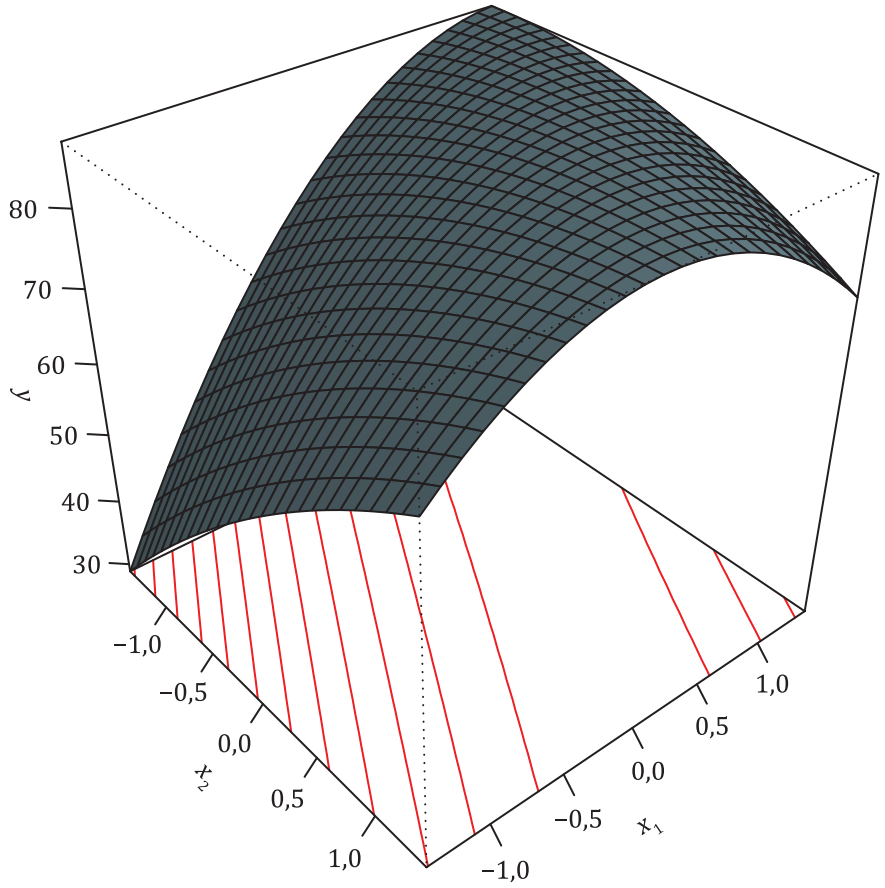
$$\hat{y} = 83{,}6 + 11{,}1x_1 + 4{,}1x_2 - 9{,}4x_1x_2 - 6{,}5x_1^2 - 0{,}4x_2^2 \quad \text{(E.6)}$$

Note that the coefficients have the same sign and are of the same order of magnitude in all four polynomials, so the four polynomials seem to be very similar. In Figures E.1 to E.4 the four second order polynomials are represented as perspective plots of the estimated response surfaces over the range $-1{,}41 < x_1 < 1{,}41$ and $-1{,}41 < x_2 < 1{,}41$. In addition, contours of constant estimated response are plotted in the x_1-x_2 plane. Plots of the response surfaces over that range would be relevant had the coefficients of the polynomials been obtained by doing experiments with the design in Table E.1.



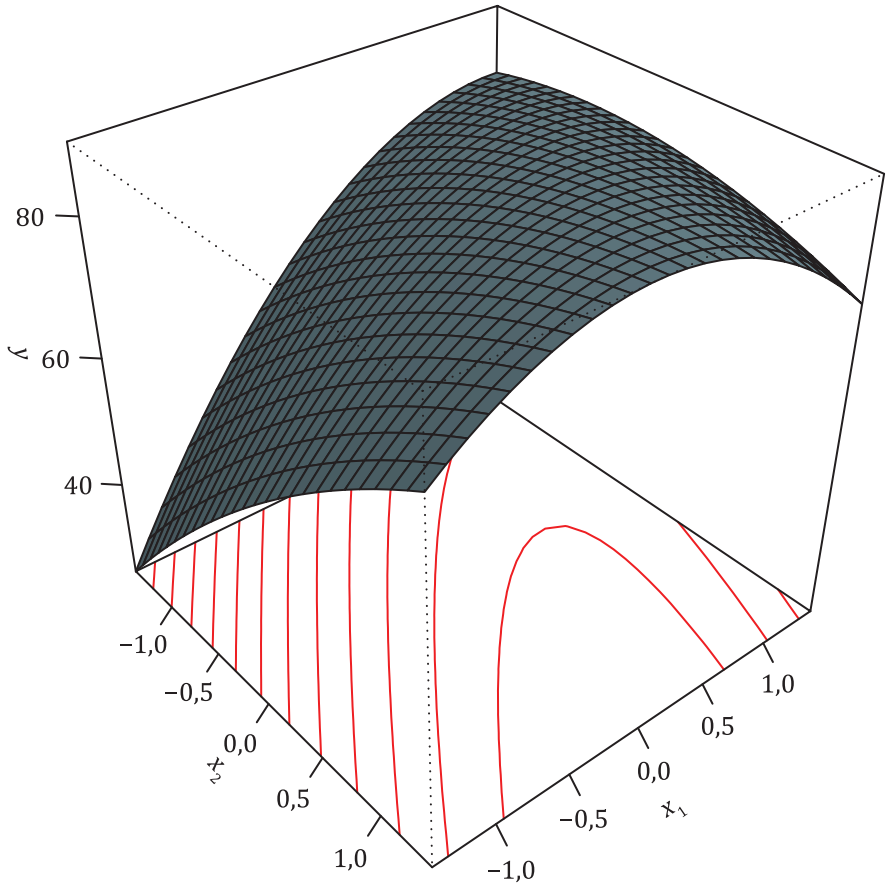
NOTE The stationary point is a maximum located inside the experimental region.

Figure E.1 — Perspective plot and contour plot of the estimated response in Formula (E.3)



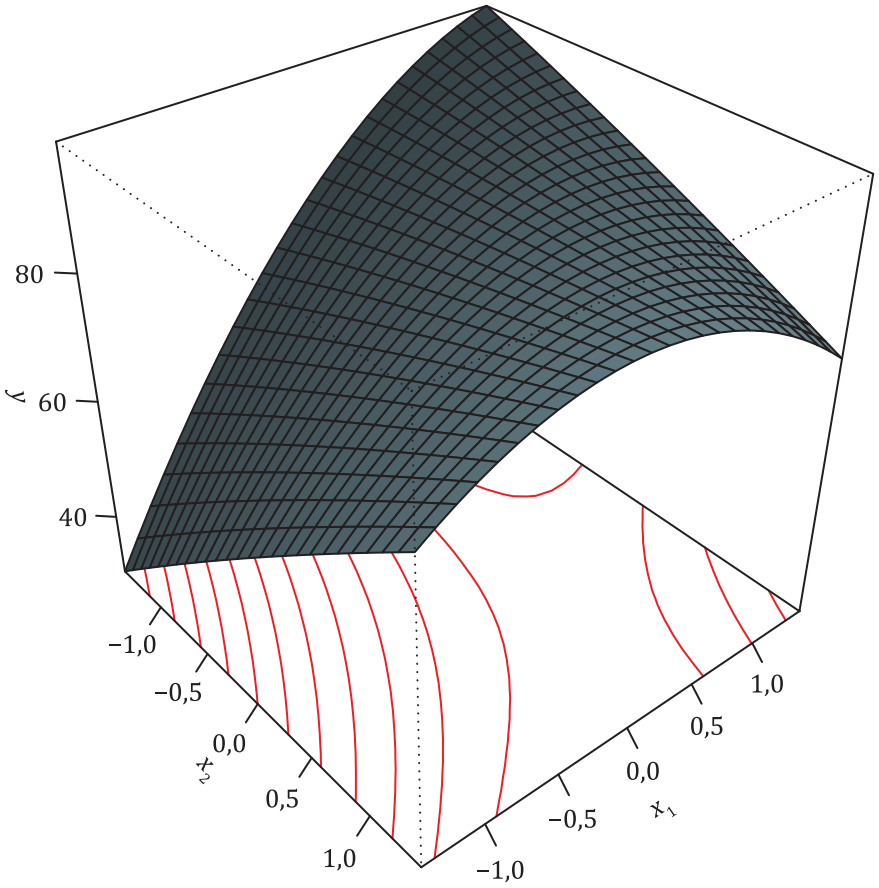
NOTE The stationary point is a saddle point located inside the experimental region. But the eigenvalues are very different so over the experimental region the response surface is a stationary ridge.

Figure E.2 — Perspective plot and contour plot of the estimated response in Formula (E.4)



NOTE The stationary point is a maximum lying outside the experimental region (x_1,x_2) = (−2,13, 5,08), so the over the experimental region the response surface is a rising ridge.

Figure E.3 — Perspective plot and contour plot of the estimated response in Formula (E.5)



NOTE The stationary point is a saddle point located inside the experimental region.

Figure E.4 — Perspective plot and contour plot of the estimated response in Formula (E.6)

E.3 Canonical analysis of response surfaces

E.3.1 Location of the stationary point

Suppose the purpose is to find the values of $x_1, x_2, \ldots, x_k$ that optimize the predicted response. This point, if it exists, will be the set of $x_1, x_2, \ldots, x_k$ for which the partial derivatives $\frac{\partial \hat{y}}{\partial x_1} = \frac{\partial \hat{y}}{\partial x_2} = \cdots = \frac{\partial \hat{y}}{\partial x_k} = 0$. This point, $\boldsymbol{x}_S$ with coordinates $x_{S1}, x_{S2}, \ldots, x_{Sk}$, is called the *stationary point*. The stationary point could represent the following:

a) a point of maximum response,

b) a point of minimum response, or

c) a saddle point.

The first and the third possibilities are shown in Figures E.1 and E.4 for the case $k = 2$.

Contour plots generated by computer software play an important role in the study of the response surface, if the number of factors is 2 or 3. By generating contour plots for a response surface under investigation, the experimenter can usually locate the optimum with reasonable precision. But a complete characterization is more difficult if only relying on contour plots. An example of the application of contour plots with three predictor variables is given in A.8.3.

A general mathematical solution for the location of the stationary point of a second order response surface can be found. Writing the second order model in matrix notation, it has the form

$$\hat{y} = b_0 + \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \tag{E.7}$$

where

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} b_{11} & \frac{b_{12}}{2} & & \frac{b_{1k}}{2} \\ \frac{b_{12}}{2} & b_{22} & \ddots & \vdots \\ \frac{b_{1k}}{2} & & \cdots & b_{kk} \end{bmatrix}. \tag{E.8}$$

That is, $\boldsymbol{b}$ is a (k×1) vector of the first order regression coefficients and $\boldsymbol{B}$ is a (k×k) symmetric matrix whose main diagonal elements are the *pure* quadratic coefficients (b_{ii}), and whose off diagonal elements are one-half the *mixed* quadratic coefficients $\frac{b_{ij}}{2}$. The derivative of $\hat{y}$ with respect to the elements of $\boldsymbol{x}$ and equated to $\mathbf{0}$ is

$$\frac{\partial \hat{y}}{\partial \boldsymbol{x}} = \boldsymbol{b} + 2\boldsymbol{B}\boldsymbol{x} = 0.$$

The stationary point is the solution to this formula or

$$\boldsymbol{x}_S = -\frac{1}{2}\boldsymbol{B}^{-1}\boldsymbol{b}. \tag{E.9}$$

Furthermore, by substituting this expression into the expression for the predicted response, the predicted response at the stationary point is

$$\hat{y}_S = b_0 + \frac{1}{2}\boldsymbol{x}_S{}'\boldsymbol{b}. \tag{E.10}$$

The distance of the stationary point to the design centre is

$$D_S = \sqrt{\sum_{i=1}^{k} x_{Si}{}^2}. \tag{E.11}$$

This distance is used to determine the location of the stationary point relative to the experimental region. If software coded variables are used it is customary to consider the experimental region as the ball with centre at the design centre and radius 1. If coded variables are used, the experimental region is the ball with centre at the design centre and radius equal to the larger of α or $\sqrt{k}$.

E.3.2 Nature of the stationary point (canonical analysis)

Canonical analysis is a method of rewriting a fitted quadratic function of the predictor variables in a form in which can be more easily understood. This is achieved by a rotation of the coordinate axes which removes all cross-product terms. This operation gives the *A-canonical form*. If desired, this may be followed by a change of origin to remove first order terms as well. The result is called the *B-canonical* form.

Let the $k \times k$ matrix $\boldsymbol{P}$ be the matrix whose columns are the normalized eigenvectors corresponding to the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ of $\boldsymbol{B}$. Then

$$\boldsymbol{P}'\boldsymbol{B}\boldsymbol{P} = \boldsymbol{\Lambda}, \tag{E.12}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{B}$. The nature of the stationary point is determined from the signs of the eigenvalues of matrix $\boldsymbol{B}$, and the relative magnitudes of these eigenvalues are helpful in further interpretation of the response surface.

The A-canonical form is obtained by rotating the coordinate system so the axes are the eigenvectors of $\boldsymbol{B}$. If the rotated coordinates are called $\boldsymbol{u}$ their relationship to the $\boldsymbol{x}$ coordinates is

$$\boldsymbol{u} = \boldsymbol{P}'\boldsymbol{x} \quad \text{or} \quad \boldsymbol{P}\boldsymbol{u} = \boldsymbol{x}.$$

Using this relationship and the fact that $\boldsymbol{P}$ is the matrix whose columns are the normalized eigenvectors so $\boldsymbol{P'P} = \boldsymbol{PP'} = \boldsymbol{I}$ the A-canonical form follows from the model Formula (E.7)

$$\begin{aligned}
\hat{y} &= b_0 + \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \\
&= b_0 + \boldsymbol{x}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{P}\boldsymbol{P}'\boldsymbol{B}\boldsymbol{P}\boldsymbol{P}'\boldsymbol{x} \\
&= b_0 + (\boldsymbol{P}'\boldsymbol{x})'\boldsymbol{P}'\boldsymbol{b} + (\boldsymbol{P}'\boldsymbol{x})'\boldsymbol{P}'\boldsymbol{B}\boldsymbol{P}(\boldsymbol{P}'\boldsymbol{x}) \\
&= b_0 + \boldsymbol{u}'\boldsymbol{a} + \boldsymbol{u}'\boldsymbol{\Lambda}\boldsymbol{u} \\
&= b_0 + \boldsymbol{u}'\boldsymbol{a} + \hat{\lambda}_1 a_1^2 + \cdots + \hat{\lambda}_k a_k^2.
\end{aligned} \tag{E.13}$$

The first order terms are useful in determining the direction of movement along the rising ridge towards the optimum.

The model Formula (E.7) can be expressed in a new coordinate system that is centred at the stationary point $\boldsymbol{x}_S$ and with coordinate axes that are parallel to the original axis, and this gives a simpler form of the model formula without a linear term. If the new coordinates are called **z**, the relation with the original coordinates **x** are

$$\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{x}_S. \tag{E.14}$$

To see this substitute $\boldsymbol{z} + \boldsymbol{x}_S$ for $\boldsymbol{x}$ in Formula (E.7)

$$\begin{aligned}
\hat{y} &= b_0 + \boldsymbol{x}'\boldsymbol{b} + \boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} \\
&= b_0 + (\boldsymbol{z} + \boldsymbol{x}_S)'\boldsymbol{b} + (\boldsymbol{z} + \boldsymbol{x}_S)'\boldsymbol{B}(\boldsymbol{z} + \boldsymbol{x}_S) \\
&= \left[b_0 + \boldsymbol{x}_S{}'\boldsymbol{b} + \boldsymbol{x}_S{}'\boldsymbol{B}\boldsymbol{x}_S\right] + \boldsymbol{z}'\boldsymbol{B} + \boldsymbol{z}'\boldsymbol{B}\boldsymbol{z} + 2\boldsymbol{z}'\boldsymbol{B}\boldsymbol{x}_S \\
&= \hat{y}_S + \boldsymbol{z}'\boldsymbol{B}\boldsymbol{z},
\end{aligned} \tag{E.15}$$

because $2\boldsymbol{z}'\boldsymbol{B}\boldsymbol{x}_S = -\boldsymbol{z}'\boldsymbol{b}$ from Formula (E.9) and the term in square brackets is $\hat{y}_S$. Rotating the coordinate system to axes that are parallel to the eigenvectors of $\boldsymbol{B}$ gives further simplification. If the rotated coordinates are called $\boldsymbol{w}$ the relationship with the $\boldsymbol{z}$ coordinates is

$$\boldsymbol{w} = \boldsymbol{P}'\boldsymbol{z} \quad \text{or} \quad \boldsymbol{P}\boldsymbol{w} = \boldsymbol{z}.$$

Substituting ***Pw*** for ***z*** in Formula (E.15) gives

$$\begin{aligned}\hat{y} &= \hat{y}_S + \boldsymbol{z}'\boldsymbol{B}\boldsymbol{z} \\ &= \hat{y}_S + \boldsymbol{w}'\boldsymbol{P}'\boldsymbol{B}\boldsymbol{P}\boldsymbol{w} \\ &= \hat{y}_S + \boldsymbol{w}'\boldsymbol{\Lambda}\boldsymbol{w} \\ &= \hat{y}_S + \hat{\lambda}_1 w_1^2 + \hat{\lambda}_2 w_2^2 + \cdots + \hat{\lambda}_k w_k^2,\end{aligned} \tag{E.16}$$

where the last step follows from Formula (E.12).The coefficients $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ are the eigenvalues of ***B*** and the variables w_1, w_2, ..., w_3 are called canonical variables. The representation in Formula (E.16) is the *B-canonical form*.

It follows from the B-canonical form in Formula (E.16) that nature of the stationary point is determined from the signs of the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ of matrix ***B***

a) If $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ are all *negative*, the stationary point is a *maximum.*

b) If $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ are all positive, the stationary point is a *minimum*.

c) If $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ have mixed signs, the stationary point is a *saddle* point.

E.3.3 Confidence intervals for the eigenvalues

It is important to realize that it is the properties of the fitted response surface that are determined by the eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ of ***B***. Although the fitted response surface may fit the underlying true surface well, it does not follow that the underlying true surface has the same properties as the fitted surface. The eigenvalues $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ of ***B*** are only estimates of the eigenvalues $\lambda_1, \ldots, \lambda_k$ of a matrix similar to ***B*** but with the estimates b_{ij} replaced by the parameters β_{ij}. Fortunately, the standard errors of $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ can be evaluated and can be used to calculate the confidence intervals for the eigenvalues of the underlying true surface. For rotatable or approximately rotatable designs the situation is simple; for the standard errors of $\hat{\lambda}_1, \ldots, \hat{\lambda}_k$ are roughly of the same size as standard error of the estimates b_{ii} of the quadratic coefficients β_{ii}.

The formula for the 95% confidence interval for λ_i is

$$\hat{\lambda}_i - se(\hat{\lambda}_i)t_{0,95}(v) \leq \lambda_i \leq \hat{\lambda}_i - se(\hat{\lambda}_i)t_{0,95}(v)$$

where $se(\hat{\lambda}_i)$ denotes the standard error of $\hat{\lambda}_i$ and v denotes the degrees of freedom for total error in the experiment.

Bibliography

[1] Box G.E.P., & Wilson K.B. On the experimental attainment of optimum conditions. J. R. Stat. Soc., B. 1951, **13** (1) pp. 1–45

[2] Box G.E.P., & Hunter J. Stuart and Hunter, W.G. (2005). *Statistics for experimenters: design, discovery, and innovation*. New York: John Wiley & Sons, Inc. 2nd edition

[3] Box G.E.P. Evolutionary operation: A method for increasing industrial productivity. Appl. Stat. 1957, **6** pp. 81–101

[4] Box G.E.P., & Draper N.R. Evolutionary Operation – A Statistical Method for Process Improvement. John Wiley & Sons, Inc, New York, 1969

[5] Box G.E.P., & Behnken D.W. Some three level designs for the study of quantitative variables. Technometrics. 1960, **2** pp. 455–476

[6] Wu and Hamada. Experiments Planning, Analysis and Parameter Design Optimization. John Wiley & Sons, Inc, New York, 2002

[7] Myers R.H., Montgomery D.C., Anderson-Cook C.M. Response Surface Methodology: Product and Process Optimization Using Designed Experiments. John Wiley & Sons, Inc, New York, 2009

[8] Hoerl R.W. Ridge Analysis 25 Years Later. Am. Stat. 1985, **39** (3) pp. 186–192

[9] Draper N.R. Ridge Analysis of Response Surfaces. Technometrics. 1963, **5** (4) pp. 469–479

[10] ISO/TR 29901, *Selected illustrations of full factorial experiments with four factors*

[11] ISO/TR 12845, *Selected illustrations of fractional factorial screening experiments*

[12] ISO/TR 12888, Selected illustrations of gauge repeatability and reproducibility studies

[13] ISO 3534-1, *Statistics — Vocabulary and symbols — Part 1: General statistical terms and terms used in probability*

[14] ISO 3534-2, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

[15] ISO 3534-3, *Statistics — Vocabulary and symbols — Part 3: Design of experiments*

ICS 03.120.30
Price based on 80 pages

This page deliberately left blank

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards -based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

Useful Contacts:

Customer Services
Tel: +44 845 086 9001
Email (orders): orders@bsigroup.com
Email (enquiries): cservices@bsigroup.com

Subscriptions
Tel: +44 845 086 9001
Email: subscriptions@bsigroup.com

Knowledge Centre
Tel: +44 20 8996 7004
Email: knowledgecentre@bsigroup.com

Copyright & Licensing
Tel: +44 20 8996 7070
Email: copyright@bsigroup.com

bsi.

...making excellence a habit.™