

This item is the archived peer-reviewed author-version of:

Optimal design of experiments for excipient compatibility studies

Reference:

Akkermans Wannes G. M., Coppenolle Hans, Goos Peter.- Optimal design of experiments for excipient compatibility studies
Chemometrics and intelligent laboratory systems - ISSN 0169-7439 - 171(2017), p. 125-139
Full text (Publisher's DOI): <https://doi.org/10.1016/J.CHEMOLAB.2017.09.012>
To cite this reference: <http://hdl.handle.net/10067/1485200151162165141>

Accepted Manuscript

Optimal design of experiments for excipient compatibility studies

Wannes G.M. Akkermans, Hans Coppenolle, Peter Goos

PII: S0169-7439(17)30017-5

DOI: [10.1016/j.chemolab.2017.09.012](https://doi.org/10.1016/j.chemolab.2017.09.012)

Reference: CHEMOM 3500

To appear in: *Chemometrics and Intelligent Laboratory Systems*

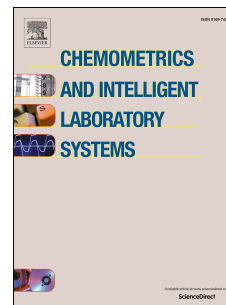
Received Date: 10 January 2017

Revised Date: 17 September 2017

Accepted Date: 19 September 2017

Please cite this article as: W.G.M. Akkermans, H. Coppenolle, P. Goos, Optimal design of experiments for excipient compatibility studies, *Chemometrics and Intelligent Laboratory Systems* (2017), doi: 10.1016/j.chemolab.2017.09.012.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Optimal design of experiments for excipient compatibility studies

Wannes G.M. Akkermans
Faculty of Bioscience Engineering
KU Leuven

Hans Coppenolle
Statistics and Decision Sciences
Janssen Pharmaceutica

Peter Goos
Faculty of Bioscience Engineering
KU Leuven, and
Faculty of Applied Economics
Universiteit Antwerpen

Abstract

A crucial stage in the development of medical drugs is to study which additives, usually called excipients, impact the active ingredient stability. This type of study is generally named an excipient compatibility study and requires a mixture experiment. Subsequently, the effect of the storage conditions, more specifically the relative humidity and temperature, on the stability is investigated. This so-called accelerated life test involves a factorial type of experiment. It has become, however, customary to include the storage conditions in the compatibility study. This provides valuable information concerning potential interactions between excipient combinations and storage conditions. Experiments that combine a mixture experiment with a factorial experiment are generally named mixture-process variable experiments. A limited number of designs for mixture process-variable experiments are available in the literature. One problem is that the proposed designs offer little flexibility. Another is that the required number of runs becomes prohibitively large for large numbers of mixture components. In this paper, we examine flexible, optimal designs for realistic mixture-process variable experiments. Our motivation is to provide guidance to pharmaceutical formulation scientists concerning state-of-the art models and designs for excipient compatibility studies. Using several proof-of-concept examples, we demonstrate that I-optimal designs offer both flexibility and small variances of prediction. We also discuss a real-life example, which could be used as a blueprint for future studies. Because many excipient compatibility studies are not completely randomized, we pay special attention to their logistics and to the resulting randomization restrictions, which lead to split-plot

and strip-plot experiments.

Keywords: D-optimal design; I-optimal design; mixture-process variable experiment; randomization restriction; split-plot design; strip-plot design.

ACCEPTED MANUSCRIPT

1 Introduction

After a new active pharmaceutical ingredient (API) has been developed, the next challenge is to select excipients for an acceptable and stable formulation. The term “stability” refers to the allowed storage time (or shelf life) of a medical drug before any degradation product achieves a level that represents a risk to the patient (Waterman and Adami, 2005), or to the capacity to retain acceptable safety and efficacy during storage. Although long regarded as being inert, excipients in the formulation of a drug can interact with the API, either directly or indirectly, by acting as catalysts for drug degradation processes (Aulton, 2013). Since excipients are an integral component of all pharmaceutical products, API stability in formulations with different excipient combinations should be assessed as early as possible in the development process to achieve the most favourable shelf life for the final drug formulation. This assessment should be performed at different temperatures and levels of humidity to gain insight into how storage conditions influence the reactions (Nyqvist, 1986).

To reduce the time-to-market (which has a major impact on the profitability), time-efficient experiments providing insights into which excipients and excipient combinations impact API stability are required. In addition to the excipient effects, also the effects of storage conditions like temperature or relative humidity are of main interest. The goal of this paper is to introduce the state-of-the-art mixture-process variable modeling techniques and experimental designs in the context of so-called excipient compatibility studies. The motivating experiment discussed in this paper involves a formulation with eight components, being the API and seven excipients. To assess the effect of the storage conditions on the stability, different temperatures and relative humidities were used in the study, following the ICH Q1A-R2 (2003) guidelines on accelerated stability studies. More specifically, in the motivating experiment, the temperature was set to 50 and 70°C and the relative humidity was set to 50% and 75%.

When optimizing a formulation, it is assumed that its characteristics only depend on the relative proportions of its components and not on its total mass. An experiment dealing with component proportions alone is called a mixture experiment (Cornell, 2002). For a

typical excipient compatibility study, however, we need a more sophisticated experiment, called a mixture-process variable experiment. This is because such a study involves experimental factors other than the component proportions too. In general, these variables are named process variables (Næs et al., 1998). In an excipient compatibility study, the factors defining the storage conditions should be regarded as process variables. Depending on how they are performed, mixture-process variable experiments may involve different kinds of grouping. When analyzing the results, the grouping structure should be accounted for. In the motivating example for this paper, several mixtures were incubated in the same climate chamber, as a result of which the experiment had a restricted randomization and correlation was present between observations from the same climate chamber. Another kind of grouping occurs when the mixtures are produced in batches and every batch is used for multiple observations.

In Section 2 of this paper, we discuss the motivating example, a typical excipient compatibility study, and its experimental setup in detail. In Section 3, we introduce general models for data from mixture experiments and describe the general form of a mixture-process variable model. In Section 4, we provide a brief overview of the literature on the design of mixture-process variable experiments. In Section 5, we describe the flexible optimal design methodology for setting up completely randomized, split-plot and strip-plot designs for mixture-process variable experiments. In Section 6, we compare various design strategies using simple proof-of-concept examples. In Section 7, we revisit the motivating example and discuss its design and the analysis of the resulting data. In Section 8, we end the paper with a general conclusion.

2 Motivating example

In our motivating example, the compatibility of an API with a number of functional excipients is tested. Seven plausible excipients are selected based on the API's chemical characteristics and desired dosage forms. Three are fillers: mannitol, micro-crystalline cellulose and lactose. The other four are so-called functional excipients: silicon dioxide as glidant, croscarmellose sodium as disintegrant, sodium lauryl sulphate as surfactant and

magnesium stearate as lubricant. Due to their specific properties, the proportions of the excipients have to satisfy the following constraints:

$$0.35 \leq m_1 + m_2 + m_3 \leq 0.86, \quad (1)$$

$$0.00 \leq m_6 \leq 0.37 \quad (2)$$

and

$$0.00 \leq m_7 \leq 0.30, \quad (3)$$

where m_1 , m_2 and m_3 are the proportions of the fillers mannitol, micro-crystalline cellulose and lactose, respectively, m_6 is the sodium lauryl sulphate proportion, and m_7 is the magnesium stearate proportion in the formulation, ignoring the API. In our motivating example, no explicit constraints were imposed on the proportions of silicon dioxide, m_4 , and croscarmellose sodium, m_5 . However, imposing a lower bound of 0.35 on the sum of the fillers' proportions implies an upper bound of 0.65 for the sum of the proportions of silicon dioxide, croscarmellose sodium, sodium lauryl sulphate and magnesium stearate.

Traditionally, to account for storage condition effects, a classic mixture experiment for the excipients is designed and this design is replicated for a limited number of storage conditions. This design strategy has two weaknesses. First, the information content of the experimental tests is suboptimal. Second, using only a limited number of storage conditions makes it impossible to draw statistical inferences concerning the effects of the storage conditions and concerning possible interactions between the storage conditions, on the one hand, and the mixture components or excipients, on the other hand.

It is generally recommended to use completely randomized designs for any kind of experiment, so that all experimental tests are independent. In our 90-run motivating example, however, six groups of 15 mixtures were incubated simultaneously in six climate chambers, so that the experiment was not completely randomized and not all experimental tests were performed independently. Consequently, the experiment involved six groups of 15

correlated observations. In the laboratory where the experiment was conducted, a robot prepared all the mixtures independently. Therefore, there was only one kind of grouping in our motivating example (due to one restriction on the randomization), resulting in a split-plot mixture-process variable experiment. In another testing facility, for lack of a robot, the mixtures are produced in batches. All samples originating from the same batch then also give rise to a group of correlated observations. The resulting mixture-process variable experiment in that facility therefore involves two kinds of grouping (one due to the incubation in the climate chambers and one due to the batches of mixtures prepared) and is called a strip-plot experiment.

During the data analysis, any correlation due to the grouping of the experimental tests should be taken into account. The importance of recognizing restrictions in randomization when conducting experiments and analyzing the resulting data in general is discussed in Goos and Jones (2011), and in Cornell (1988) and Kowalski et al. (2002) for mixture-process variable experiments.

3 Statistical model

Statistical models for data from mixture and mixture-process variable experiments differ from the ones traditionally used for factorial experiments. The same goes for models for data from excipient compatibility studies. This is due to the fact that, when dealing with a mixture containing q components with proportions m_1, m_2, \dots, m_q , all test combinations satisfy the bounds

$$0 \leq m_i \leq 1$$

and the general mixture constraint

$$\sum_{i=1}^q m_i = 1. \quad (4)$$

In words, the proportion of each component in the formulation is bounded between zero and one, and the sum of all component proportions in a specific formulation is one. After the dose-ranging study in phase I clinical trials, the proportion of the API in the final

formulation is determined and remains fixed during the whole excipient compatibility study. For this reason, the API's proportion is generally not considered explicitly in excipient compatibility studies (e.g. Piepel, 1999; Anderson-Cook et al., 2004; Martinello et al., 2006), and the remaining components' proportions are rescaled so that they sum to one (Cornell, 2002). A common statistical model for data from a mixture experiment is the so-called 2nd-order canonical polynomial, proposed by Scheffé (1958):

$$E(Y) = \sum_{i=1}^q \beta_i m_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} m_i m_j, \quad (5)$$

where β_i represents the expected response for a formulation containing only component i , and β_{ij} is proportional to the difference between the expected response when $m_i = m_j = 0.5$ and the average expected responses at $m_i = 1$ and $m_j = 1$ (Cornell, 2002). The first kind of parameter is called a “pure blending” parameter, while the latter is a “binary blending” parameter. Interpreting the parameters β_i and β_{ij} is quite complicated in mixture experiments, because increasing one proportion requires decreasing at least one other proportion. Therefore, the model parameters cannot be interpreted independently, unlike in traditional regression models. A more elaborate discussion on models for mixture experiments can be found in Cornell (2002) and Smith (2005).

In an excipient compatibility study, the effects of additional factors related to the storage, such as temperature and relative humidity, are also of primary interest to the researchers. These additional factors are generally referred to as process variables. As a result, excipient compatibility studies typically belong to the class of mixture-process variable experiments. The first generation of mixture-process variable models have been proposed by Cornell and Gorman (1984) and Cornell (1988). However, these models become impractically large when many mixture components or process variables are to be studied. For this reason, Kowalski, Cornell and Vining (2000) proposed a new generation of mixture-process variable models which involve fewer parameters, and which therefore require fewer data points and allow for smaller experiments.

While the first generation of mixture-process variable models simply crossed a 2nd-order

Scheffé model with a traditional response surface model in the process variables, Kowalski et al. (2000) started with the 2nd-order Scheffé model in Equation (5) and the 2nd-order response surface model

$$E(Y) = \alpha_0 + \sum_{i=1}^r \alpha_i p_i + \sum_{i=1}^{r-1} \sum_{j=i+1}^r \alpha_{ij} p_i p_j + \sum_{i=1}^r \alpha_{ii} p_i^2, \quad (6)$$

where p_1, p_2, \dots, p_r represent the levels of r process variables. They combined both models in such a way that the final model only includes first- and second-order terms. This resulted in the mixture-process variable model

$$E(Y) = \sum_{i=1}^q \beta_i m_i + \sum_{i=1}^{q-1} \sum_{j=i+1}^q \beta_{ij} m_i m_j + \sum_{i=1}^q \sum_{j=1}^r \gamma_{ij} m_i p_j + \sum_{i=1}^{r-1} \sum_{j=i+1}^r \alpha_{ij} p_i p_j + \sum_{i=1}^r \alpha_{ii} p_i^2. \quad (7)$$

That model involves the main effects of all component proportions, all two-way interactions involving the component proportions and/or the process variables, and the quadratic effects of all process variables. The model does not involve the main effects of the process variables, because these are inestimable when all interactions between the component proportions and the process variables are included in the model (for a technical explanation, see Smith (2005) and Goos and Jones (2011)).

The mixture-process variable model in Equation (7) is parsimonious because it involves only 1st- and 2nd-order terms. The model implicitly assumes that interaction effects of any two mixture components do not depend on the levels of the process variables. To relax that assumption, it may be useful to add the following terms to the model:

$$\sum_{i=1}^{q-1} \sum_{j=i+1}^q \sum_{k=1}^r \gamma_{ijk} m_i m_j p_k. \quad (8)$$

As excipient compatibility studies often involve many experimental tests, we recommend to evaluate these terms and check whether they add value to the model. The additional terms describe the so-called interactions between the process variables and the non-linear blending terms.

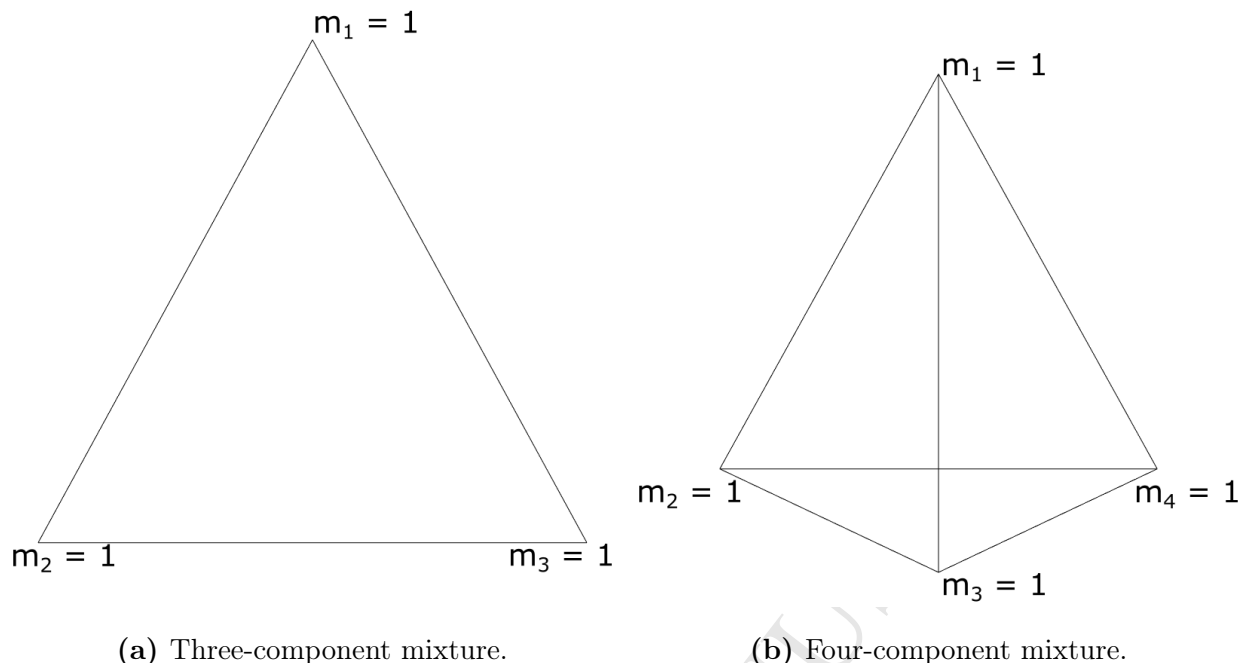


Figure 1: Experimental regions of unconstrained mixture experiments.

4 Designing mixture-process variable experiments

4.1 Experimental region for component proportions

When there are no bounds or constraints on the mixture component proportions other than the general mixture constraint, the experimental region is called a “regular simplex”. Due to the general mixture constraint in Equation (4), we can express any one component proportion as a linear function of the other component proportions:

$$m_q = 1 - \sum_{i=1}^{q-1} m_i. \quad (9)$$

As a result, we can visualize the experimental region of mixture experiments with up to four components graphically, as shown in Figure 1. For three components, the experimental region is an equilateral triangle. For four components, it becomes a tetrahedron.

The edge points, or vertices, in Figure 1 correspond to formulations with a proportion of 1 for one component. So, the vertices correspond to pure components. When moving from one vertex to another along an edge of the simplex, the proportion of the component

corresponding to the initial vertex is reduced, while the proportion of the component corresponding to the other vertex increases, so as to satisfy the mixture constraint. The edges of the simplex thus correspond to binary mixtures, comprising only two components. Points within a simplex, sometimes called interior points, always represents a feasible mixture with a specific non-zero proportion for each component.

When multicomponent constraints on the proportions are present, the experimental region is generally not a regular simplex any more. An example of a multicomponent constraint is given by Equation (1). When many constraints are present, the experimental region becomes an irregularly shaped volume within the simplex.

4.2 Mixture designs

For problems where the experimental region is a regular simplex, standard mixture designs have been proposed. The best-known standard designs are the simplex-lattice and simplex-centroid designs introduced by Scheffé (1958, 1963) and the designs of Draper and Lawrence (1965a,b). When the experimental region is not a regular simplex, it becomes much more difficult to construct a suitable experimental design. Using the programs of Piepel (1988), one could construct a so-called extreme vertex design. An extreme vertex design uses all vertices of the experimental region. An extreme vertex design of degree 2 also involves the edge mid-points, i.e., the averages of all pairs of vertices. An extreme vertex design of degree n also involves the averages of all n -tuples of vertices. The number of runs of these kinds of designs increase rapidly with the number of mixture components and with the degree. For example, for the motivating example, the extreme vertex design of degree 3 involves 696 experimental runs, making the design infeasibly large.

In those cases, it is more sensible to use computer-aided optimal design of experiments to generate a feasible, informative design with the desired number of runs. Optimal designs are especially useful when the experimental region is restricted or when only a limited number of runs is possible. A common criticism on optimal experimental designs is that they are model-dependent and no longer optimal when the model is changed during the

analysis, e.g. by dropping insignificant terms. However, standard mixture designs also have been created with a certain model in mind and also become “suboptimal” when estimating a different model. One approach to compute optimal experimental designs is to use a point-exchange algorithm that uses the extreme vertices as well as their pairwise averages and, possibly, other linear combinations of the vertices as a candidate set. The algorithm then selects the best candidates according to some optimality criterion. Another approach is to use a coordinate-exchange algorithm which starts from a random design. This kind of algorithm finds an optimal design by changing the coordinates of the random starting design one at a time. Examples of such algorithms for mixture experiments can be found in Piepel et al. (2005) and Ruseckaite et al. (2017). Coordinate-exchange algorithms have the advantage that all points in the experimental region can be selected, which increases the probability of finding the globally optimal experimental design (Goos et al., 2016).

4.3 Mixture-process variable designs

In general, a completely randomized design is recommended for any kind of experiment. The first completely randomized design for a mixture-process variable experiment was proposed by Scheffé (1963). For an example with three mixture components and one process variable, Scheffé suggested to run a simplex-centroid design at all process variable levels. This results in a so-called crossed design. Using this approach, the number of experimental runs increases very rapidly with the number of process variables. For dealing with multiple process variables, Cornell and Gorman (1984) therefore suggested to use only a fraction of the level combinations of the process variables. They proposed two types of designs. The first type, called a matched fraction, places the same fractional factorial design at each simplex-centroid design point. The second type, called a mixed fraction, uses different fractional factorial designs for the process variables at each simplex-centroid design point. Using this approach, Cornell and Gorman (1984) reduce the number of observations from 56 to 14 for one example involving three mixture components and three process variables. Of course, using this method, several terms from the complete mixture-process variable model are not estimable any more.

Together with their new generation of mixture-process variable models, Kowalski et al. (2000) also proposed a new design methodology. Their main idea was to also use only a fraction of the runs from the complete crossed design, but they suggested to fractionate the simplex-centroid design rather than the factorial design. They were able to construct a design with only 24 runs for estimating all 15 parameters in their mixture-process variable model for three mixture components and two process variables. We use the resulting design, called “design F”, as a benchmark design when discussing our completely randomized proof-of-concept example.

In many experiments, including excipient compatibility studies, it is impractical or even infeasible to conduct a completely randomized experiment. When the randomization is restricted so that the runs composing the experiment appear in groups, the assumption of independence is violated. Depending on the exact logistics of the experiment, the experimental design either is a split-plot design or a strip-plot design. In our motivating example, for instance, 15 mixtures are incubated in each of six climate chambers. As a result, the experiment involves six groups of 15 correlated observations. Since the mixtures were prepared independently for each experimental run, this is the only grouping present and the experimental design is a split-plot design. Kowalski, Cornell and Vining (2002) adapted their original completely randomized designs to cope with split-plot scenarios. A second type of grouping occurs whenever multiple samples are taken from a limited number of batches. In the event the two kinds of grouping occur simultaneously, the experimental design is a strip-plot design. In the literature, there are no ready-to-use strip-plot mixture-process variable designs.

The main drawback of the designs proposed by Kowalski et al. (2000, 2002) is that they are based on simplex-centroid designs. Therefore, they are only feasible when the experimental region for the mixture component proportions is a regular simplex. One solution to cope with irregularly shaped experimental regions is to use an extreme vertex design. Another problem arises when the feasible total number of runs and/or the feasible number of runs per group differ from those of the completely randomized or split-plot mixture-process vari-

able designs of Kowalski et al. (2000, 2002). This may happen when the total budget is limited or when the total number of mixtures in each climate chamber is limited, or when only a small number of different batches can be produced. It is unclear how to modify the proposed designs in these cases.

If a completely randomized mixture-process variable experiment is feasible and no additional grouping structure is present, one could opt for the same kind of computer-aided optimal experimental design strategy as with ordinary mixture designs. One can either use a point-exchange or a coordinate-exchange algorithm. When one type of grouping is present, the point-exchange algorithm proposed by Goos and Donev (2007) to generate optimal split-plot designs provides complete flexibility in the number of runs for each group, the number of groups in the experiment, the model under consideration and the bounds on the mixture component proportions and the process variables. When two types of grouping are present, the coordinate-exchange algorithms of Arnouts et al. (2010, 2013) can be used. In the next section, the optimal experimental design approach is discussed in detail.

5 Optimal design of experiments

In order to construct optimal experimental designs, it is important to start from an *a priori* model and to know which estimation technique will be used for that model. For this reason, we start this section by introducing the model-expansion function, the model matrix and the ordinary and generalized least squares estimators. We explain that the required estimator depends on whether grouping is present in the experimental runs or not.

5.1 Model estimation

For a good understanding of optimal experimental design, we first need a function, say \mathbf{f} , that maps the design points (i.e., the combinations of component proportions and process variable levels tested in the experiment) from the experimental region χ onto the model space $\mathbf{f}(\chi)$. For example, the mapping function for a mixture-process variable model for

an experiment with q mixture components and r process variables is

$$\mathbf{f} : \mathbb{R}^{(q+r)} \rightarrow \mathbb{R}^s : \mathbf{x}_i \rightarrow \mathbf{f}(\mathbf{x}_i), \text{ for } i = 1, \dots, n, \quad (10)$$

where n is the total number of experimental runs, s is the number of parameters in the full model and \mathbf{x}_i is the mixture-process variable setting for the i^{th} run. For instance, for a mixture-process variable experiment with three mixture components ($q = 3$) and two process variables ($r = 2$), the mapping function for a run i involving the pure component 1 and the process variables' low level (denoted by $m_1 = 1$, $m_2 = m_3 = 0$ and $p_1 = p_2 = -1$) is

$$\begin{aligned} \mathbf{x}_i &= \begin{bmatrix} m_1 & m_2 & m_3 & p_1 & p_2 \\ 1 & 0 & 0 & -1 & -1 \end{bmatrix} \rightarrow \\ \mathbf{f}(\mathbf{x}_i) &= \begin{bmatrix} m_1 & m_2 & m_3 & m_1 m_2 & m_1 m_3 & m_2 m_3 & m_1 p_1 & m_1 p_2 & m_2 p_1 & m_2 p_2 & m_3 p_1 & m_3 p_2 & p_1 p_2 & p_1^2 & p_2^2 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

Stacking all vectors $\mathbf{f}(\mathbf{x}_i)$ yields the $(n \times s)$ matrix \mathbf{X} , which describes the experiment in the model space and which is called the model matrix. In the remainder of this paper, the function \mathbf{f} is called the model-expansion function. The least squares method is usually used for estimating the parameters in the statistical model. In general, the least squares estimator is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}, \quad (11)$$

where \mathbf{Y} is a vector containing the n responses corresponding to the rows of \mathbf{X} and \mathbf{V} is the $(n \times n)$ variance-covariance matrix of the responses.

When dealing with a completely randomized experiment, all the experimental runs are independent and the responses are uncorrelated. So, for a completely randomized experiment, $\mathbf{V} = \sigma_\epsilon^2 \mathbf{I}_n$, with σ_ϵ^2 describing the run-to-run variation and \mathbf{I}_n the $(n \times n)$ identity matrix. In that case, the least squares estimates are called ‘‘ordinary least squares’’ estimates. When grouping is present in the data, i.e., when we are dealing with a split-plot or strip-plot experiment, the experimental runs are no longer independent, their responses are correlated, and $\mathbf{V} \neq \sigma_\epsilon^2 \mathbf{I}_n$. In such cases, the least squares estimates are called ‘‘generalized

least squares” estimates.

Split-plot and strip-plot experiments find their origin in agricultural experimentation. Figure 2a shows a simple two-factor split-plot design. A split-plot experiment generally involves multiple plots or fields, traditionally called whole plots. In Figure 2a, these whole plots are represented by the columns. A level of the whole-plot or hard-to-change factor is allocated randomly to each of the columns, e.g. irrigation with levels I_1 and I_2 . The whole plots are then divided in sub-plots, and a level of a sub-plot or an easy-to-change factor, e.g. fertilizer with levels F_1 and F_2 , is allocated randomly to each sub-plot. Due to the fact that all the observations from within one whole plot are performed on the same field, their responses are correlated. So, in an agricultural split-plot design, the grouping is caused by the different fields or whole plots.

In a strip-plot experiment, as shown in Figure 2b, any field used is partitionated in a number of horizontal and a number of vertical strips. This specific design is used when there are two hard-to-change factors. The first hard-to-change factor, applied to the vertical strips, is again irrigation. Unlike in the split-plot case, the fertilizer cannot be applied randomly within the vertical strip, but has to be applied to horizontal strips over the entire field. Thus, the fertilizer becomes the second hard-to-change factor. It should be clear that not only the observations within one vertical strip are correlated, but also the observations within one horizontal strip. When an experimenter is not bounded to a physical “field”, it is not necessary to explore all row-column combinations.

In a split-plot experiment with groups of k observations, the k responses within a group are correlated, and the $(k \times k)$ variance-covariance matrix of these k responses is

$$\mathbf{\Lambda} = \begin{bmatrix} \sigma_\lambda^2 + \sigma_\epsilon^2 & \sigma_\lambda^2 & \cdots & \sigma_\lambda^2 \\ \sigma_\lambda^2 & \sigma_\lambda^2 + \sigma_\epsilon^2 & \cdots & \sigma_\lambda^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\lambda^2 & \sigma_\lambda^2 & \cdots & \sigma_\lambda^2 + \sigma_\epsilon^2 \end{bmatrix}.$$

In this matrix, σ_ϵ^2 again represents the run-to-run variation in the responses, while σ_λ^2

$I_2 F_1$	$I_1 F_1$	$I_2 F_2$	$I_1 F_1$	$I_2 F_2$	$I_1 F_2$	$I_2 F_2$	$I_1 F_2$
$I_2 F_1$	$I_1 F_2$	$I_2 F_1$	$I_1 F_2$	$I_2 F_1$	$I_1 F_1$	$I_2 F_1$	$I_1 F_1$
$I_2 F_2$	$I_1 F_1$	$I_2 F_1$	$I_1 F_1$	$I_2 F_2$	$I_1 F_2$	$I_2 F_2$	$I_1 F_2$
$I_2 F_2$	$I_1 F_2$	$I_2 F_2$	$I_1 F_2$	$I_2 F_1$	$I_1 F_1$	$I_2 F_1$	$I_1 F_1$

(a) Split-plot experiment.

(b) Strip-plot experiment.

Figure 2: Simple split- and strip-plot experiment examples.

represents the group-to-group variation. In the motivating example, the climate chambers actually used in the excipient compatibility study can be viewed as a random sample from an entire population of climate chambers. It is assumed that each climate chamber has a different effect on the response. More specifically, it is assumed that the responses for all the runs in some climate chambers are systematically higher than in the ‘average’ climate chamber, and that the responses for all the runs in some other climate chambers are systematically lower than in the ‘average’ climate chamber, due to chance. How much higher and lower depends on the application. In some applications, σ_λ^2 is large relative to σ_ϵ^2 , in which case the shift in responses due to the different climate chambers is large compared to the run-to-run variation. In other applications, σ_λ^2 is small relative to σ_ϵ^2 , in which case the shift in responses due to the different climate chambers is small compared to the run-to-run variation. In the jargon, we say that each climate chamber has a different random effect, which is assumed to be randomly drawn from a normal distribution with zero mean and variance σ_λ^2 . All observations within the same climate chamber are subject to the same random effect, which is why they produce correlated responses. By making a few technical assumptions, it can be shown that, when dealing with a split-plot experiment

involving n/k groups of k correlated responses, the complete variance-covariance matrix \mathbf{V} has the following structure:

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

where $\mathbf{0}_{k \times k}$ is a $(k \times k)$ matrix of zeros.

When dealing with a strip-plot experiment, both the grouping in rows and the grouping in columns result in sets of correlated observations. Without using matrix algebra, it is more complicated to write down a general structure for \mathbf{V} when a strip-plot design is used than when a split-plot design is used. The general idea remains the same, but there are two kinds of random effects, one for the first kind of grouping and one for the second kind. So, when using a strip-plot experimental design, there are three variances: σ_{ϵ}^2 for the run-to-run variation, σ_{λ}^2 to describe variation between groups of the first kind, and σ_{μ}^2 to describe variation between groups of the second kind. An instructive example of a \mathbf{V} matrix for a strip-plot design is given in Chapter 11 of Goos and Jones (2011).

In practice, the \mathbf{V} -matrix has to be estimated from the data. Letsinger et al. (1996), Goos et al. (2006) and Langhans et al. (2005) recommend the restricted maximum likelihood (REML) methodology to this end. For constructing confidence intervals and performing significance tests, it is best to use the corrected degrees of freedom and standard errors discussed in Kenward and Roger (1997, 2009).

Regardless of whether the experiment is conducted using a completely randomized design, a split-plot design or a strip-plot design, the $(s \times s)$ variance-covariance matrix of the model parameter estimates is

$$\text{var}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Sigma} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}.$$

That matrix describes the uncertainty about the model parameters after the experiment has

been conducted. It plays an important role in determining confidence intervals, performing significance tests and calculating prediction variances.

5.2 Optimality criteria

The criterion used to select an experimental design should reflect the goals of the experimenter. The most commonly used criteria are estimation-oriented and prediction-oriented. As pointed out by Shmueli (2010), who contrasts predictive and explanatory modeling, even to professional statisticians, it is not always clear what the aim of a study is. However, in the context of industrial experimentation, Montgomery (2012) points out the utility of prediction-oriented designs for response surface exploration by writing that prediction-oriented criteria “would be most likely used for second-order models, as second-order models are often used for optimization, and good prediction properties are essential for optimization.” So, the use of prediction-oriented design criteria in response surface settings is recognized as appropriate. For mixture experiments, which can be viewed as special types of response surface experiments, Scheffé (1958) already suggested using the variance of the predicted response as a starting point to design mixture experiments. Therefore, for excipient compatibility experiments, using an optimality criterion focusing on precise predictions seems most appropriate. Because the best-known criterion for selecting designs is the estimation-oriented D-optimality criterion, however, we will compare D-optimal experimental designs with prediction-oriented I-optimal designs throughout this article.

5.2.1 D-optimality

The D-optimality criterion is the most commonly used criterion for selecting a design. A D-optimal design minimizes the determinant of the variance-covariance matrix Σ of the model parameter estimates. In doing so, a D-optimal design, in a way, minimizes the uncertainty about the model parameters that will be left after the experimental data have been analyzed. It can be shown that a D-optimal design minimizes the volume of the confidence ellipsoid about the model parameters. A measure to compare the quality of different designs for a given statistical model is the relative D-efficiency. For two designs

with variance-covariance matrices Σ_1 and Σ_2 , the relative D-efficiency is defined as follows:

$$\text{Relative D-efficiency of design 1 to design 2} = \left(\frac{|\Sigma_2|}{|\Sigma_1|} \right)^{1/s}. \quad (12)$$

When the relative D-efficiency is larger than one, it indicates that design 1 is better than design 2 in terms of D-optimality.

5.2.2 I-optimality

In many cases, the main focus of mixture-process variable experiments is on the optimization of the formulation. This requires good prediction properties from the estimated model. Therefore, when selecting an experimental design, it seems reasonable to focus on prediction properties. Two optimality criteria are commonly used when focusing on prediction properties: G-optimality and I-optimality. While G-optimality seeks designs that minimize the maximum prediction variance over the experimental region, I-optimality seeks designs that minimize the average prediction variance over the experimental region. Rodríguez et al. (2010) suggest to use I-optimality rather than G-optimality, because prediction variances produced by G-optimal experimental designs are larger than those produced by I-optimal designs over most of the experimental region. Therefore, in the recent literature about response surface designs, the focus is on I-optimal designs rather than on G-optimal designs (Jones and Goos, 2012, 2015; Goos and Syafitri, 2014; Arnouts and Goos, 2015; Syafitri et al., 2015; Goos et al., 2016).

Evaluating the predictive quality of two designs in terms of I-optimality can be done by using the relative I-efficiency:

$$\text{Relative I-efficiency of design 1 to design 2} = \frac{\text{Average prediction variance of design 2}}{\text{Average prediction variance of design 1}}. \quad (13)$$

When this efficiency is larger than one, it indicates that design 1 is better in terms of I-optimality, and thus has a lower average variance of prediction than design 2.

Goos et al. (2016) showed that, when comparing I- and D-optimal mixture designs, I-optimal designs perform reasonably well in terms of D-optimality, while the performance of D-optimal designs in terms of I-optimality strongly depends on which design points are replicated and may be rather poor. This provides an additional argument for using an I-optimal design for mixture-process variable experiments rather than a D-optimal design. In the next section, we compare D- and I-optimal completely randomized, split-plot and strip-plot designs for mixture-process variable experiments to alternatives from the literature. This kind of comparison is new to the literature, as is the study of I-optimal mixture-process variable designs.

6 Proof-of-concept examples

In this section, we discuss three proof-of-concept examples. The first example involves a completely randomized experiment. Secondly, we look at a more realistic problem involving one kind of grouping of the observations. More specifically, we consider a split-plot mixture-process variable experiment. Finally, we look at a design problem in which two types of grouping occur simultaneously. This results in a strip-plot experiment.

All three examples involve three mixture components, with proportions denoted by m_1 , m_2 , and m_3 , and two process variables, with levels denoted by p_1 and p_2 . For the completely randomized experiment, we choose an experimental budget of 24 runs, because that number allows us to use the completely randomized design proposed by Kowalski et al. (2000) as a benchmark. For the case of a split-plot mixture-process variable experiment, we extend the budget to 40 runs, equally distributed over 10 groups. This allows us to use the 40-run split-plot design proposed by Kowalski et al. (2002) as a benchmark design. For the case of a strip-plot experiment, no benchmark designs can be found in the literature. Therefore, for this case, we only consider a D-optimal design and an I-optimal design.

The second-order mixture-process variable model for three mixture components and two

process variables proposed by Kowalski et al. (2000) contains 15 parameters:

$$E(Y) = \sum_{i=1}^3 \beta_i m_i + \sum_{i=1}^2 \sum_{j=i+1}^3 \beta_{ij} m_i m_j + \sum_{i=1}^3 \sum_{k=1}^2 \gamma_{ik} m_i p_k + \alpha_{12} p_1 p_2 + \sum_{i=1}^2 \alpha_{ii} p_i^2. \quad (14)$$

The first six terms of that model form the second-order Scheffé mixture model. The next six terms involve two-factor interactions between the three mixture component proportions and the two process variables. The final three terms correspond to the two-factor interaction effect of the two process variables and their quadratic effects.

The designs proposed by Kowalski et al. (2000, 2002) all share similar important features. Firstly, a classic simplex-centroid design is obtained when the design is projected across the levels of any process variable. This feature is attractive when one of the process variables is found to be unimportant. Secondly, the designs involve ternary mixtures and therefore explore the center of the mixture region ($m_1 = m_2 = m_3 = 1/3$). This makes the estimation of a ternary blending parameter possible. Thirdly, one or more center points ($m_1 = m_2 = m_3 = 1/3, p_1 = p_2 = 0$) are replicated. These replicates can be used to estimate the pure error. This pure error estimation makes it possible to formally test for lack of fit. The major downside of the benchmark designs is that the center runs ($m_1 = m_2 = m_3 = 1/3$) at the various process variable settings do not contribute much to the estimation of the parameters of interest. In addition, it is not straightforward how to adapt the benchmark designs when restrictions on the mixture component proportions are present or when a different number of runs is required.

6.1 Completely Randomized Designs

We start by comparing the 24-run benchmark completely randomized design, with both a D-optimal and an I-optimal design. The two optimal designs were generated with the mixture coordinate-exchange algorithm from JMP 12 using 1000 random starts. The three designs are shown in Figures 3, 4 and 5. In the figures, the larger bullets represent replicated design points. The large $-1, 0$ and 1 values at the top and right-hand sides of the figures represents the low, middle and high values of the two process variables.

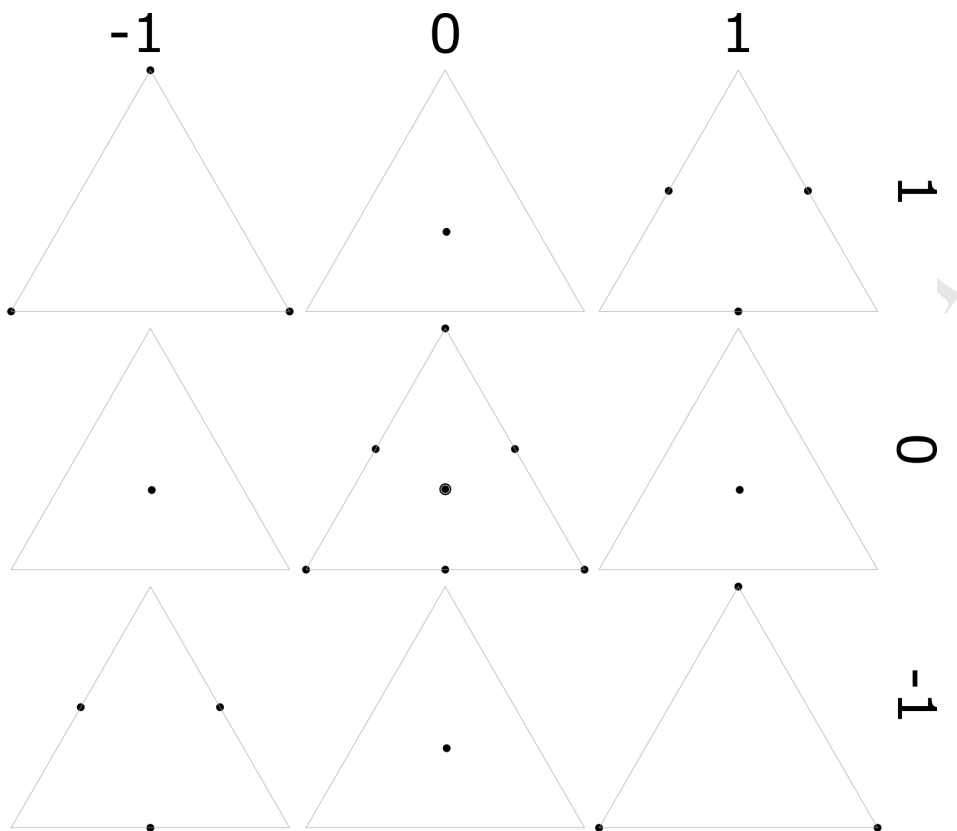


Figure 3: Benchmark completely randomized design.

The model-driven nature of D-optimal designs becomes clear when looking at Figure 4. Focusing on the precise estimation of the parameters in the model in Equation (14), the D-optimal design does not involve any mixtures with non-zero proportions for all three components at the same time. This is due to the fact that there is no ternary blending term in the model. We obtain a $\{3, 2\}$ simplex-lattice design when the D-optimal mixture-process variable design is projected across the low and high levels of any process variable. The $\{3, 2\}$ simplex-lattice design incorporates the three vertices and all the edge mid-points. At least four of the six runs of the $\{3, 2\}$ simplex-lattice design are replicated in the projections. Because there are no replicated runs in the full design, it is impossible to obtain a pure-error estimate of the residual error variance σ_e^2 from the D-optimal design. As a consequence, lack of fit cannot be formally tested.

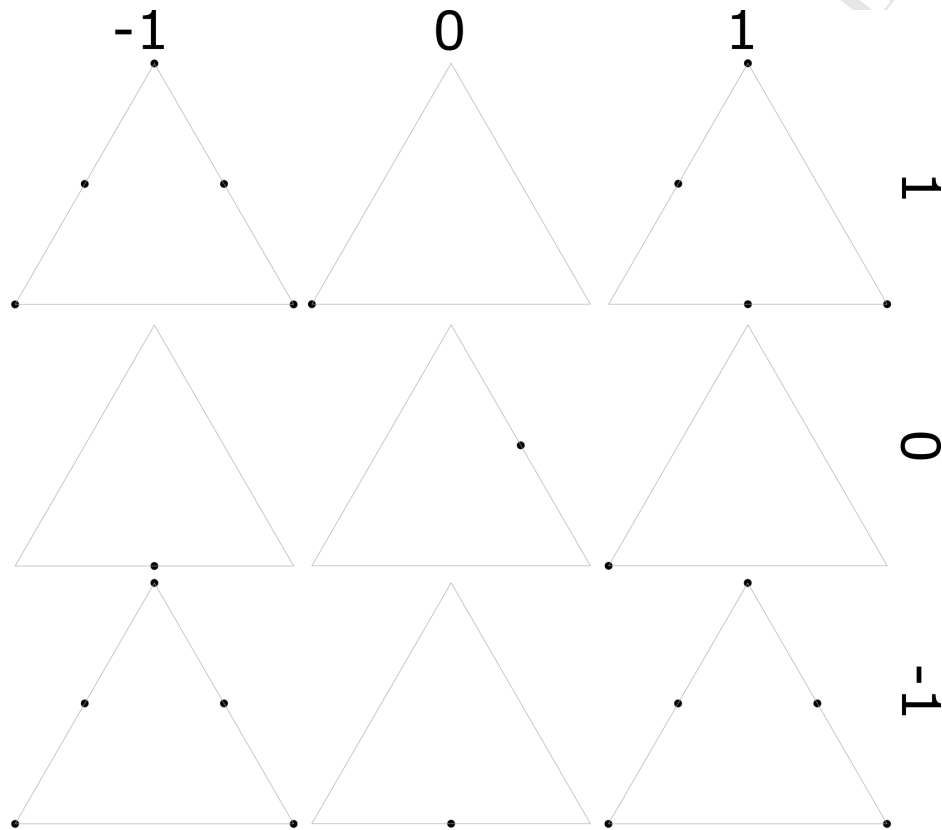


Figure 4: D-optimal completely randomized design.

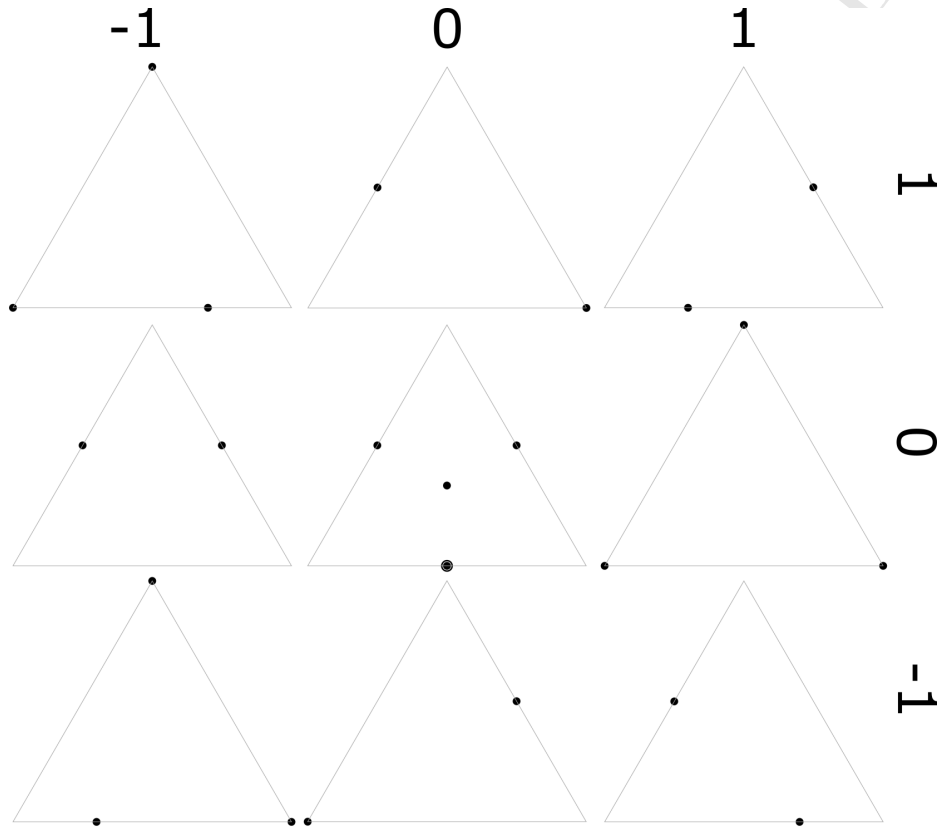


Figure 5: I-optimal completely randomized design.

Table 1: Diagnostics and relative efficiencies of the three completely randomized designs under comparison.

	$ \Sigma $	Max var	Avg var		Rel D-eff	Rel I-eff
Benchmark	5.69×10^{-5}	2.1484	0.3563	I-optimal to D-optimal	0.87	1.73
D-optimal	5.81×10^{-7}	1.1166	0.5172	Benchmark to D-optimal	0.74	1.45
I-optimal	4.44×10^{-6}	1.4017	0.2981	Benchmark to I-optimal	0.84	0.84

For I-optimal designs, the focus lies on a precise prediction, rather than on precise estimation of the parameters. Unlike the D-optimal design, the I-optimal design in Figure 5 does explore the center of the component proportions' simplex-shaped experimental region. This lowers the average variance of prediction. As pointed out by Goos et al. (2016), in I-optimal mixture designs for second-order Scheffé models, the binary blends have a higher weight than the pure components and the ternary blends are included as well. This is different from D-optimal designs, in which pure components and binary blends are equally important and ternary blends are not used. In the I-optimal design in Figure 5, there are 14 runs involving binary blends and only nine involving pure components. At $p_1 = p_2 = 0$, one design point is replicated, making it possible to estimate pure error and formally test for lack of fit.

The relative efficiencies and diagnostics of the benchmark design, the D-optimal design and the I-optimal design, assuming that the error variance is one, are listed in Table 1. Both the benchmark and the I-optimal design have a smaller average variance of prediction than the D-optimal design, while the D-optimal design's maximum variance of prediction is smallest. The benchmark design has the largest maximum prediction variance. In terms of D-optimality, both the D- and I-optimal design outperform the benchmark design. Both the I-optimal design and the benchmark design substantially outperform the D-optimal design in terms of I-optimality.

The Fraction of Design Space (FDS) plot in Figure 6 shows the prediction variances, ranked from small to large, for 10,000 randomly chosen points in the experimental region, for each of the designs under comparison. The plot shows that, in terms of variance of prediction, there is some difference between the I-optimal design and the benchmark design in favor of the I-optimal design. Both designs produce a smaller variance of prediction across almost

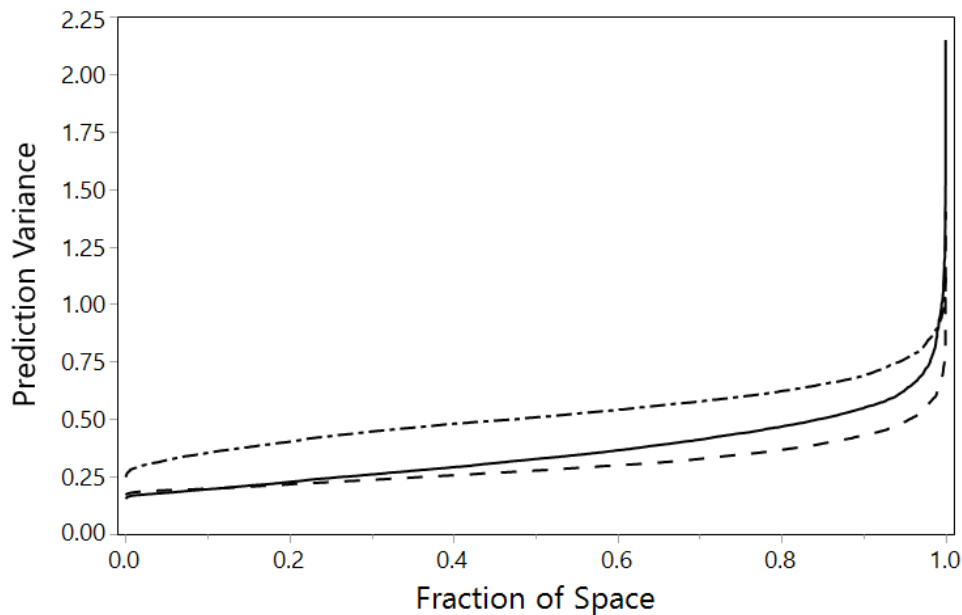


Figure 6: Fraction of Design Space plot for the benchmark (solid), D-optimal (dash-dotted) and I-optimal (dashed) completely randomized designs.

the entire experimental region than the D-optimal design.

In conclusion, the I-optimal design represents the best trade-off between good predictive capabilities and sufficient estimation precision. In addition, the *a priori* selected model to construct both the D- and I-optimal design can be extended with higher order interaction terms between the process variables and non-linear blending terms. The possibility to estimate pure error and formally test for lack-of-fit is another advantage offered by the I-optimal design.

6.2 Split-plot designs

As pointed out by Chipman et al. (1997), Bisgaard (2000), and Jones and Goos (2007), industrial experiments are often conducted using a split-plot design. This is especially true for mixture-process variable experiments (see, for instance, Cornell (1988), Kowalski et al. (2002), and Goos and Donev (2007)). For our proof-of-concept split-plot example, we introduce grouping at the level of the process variables. In our motivating example,

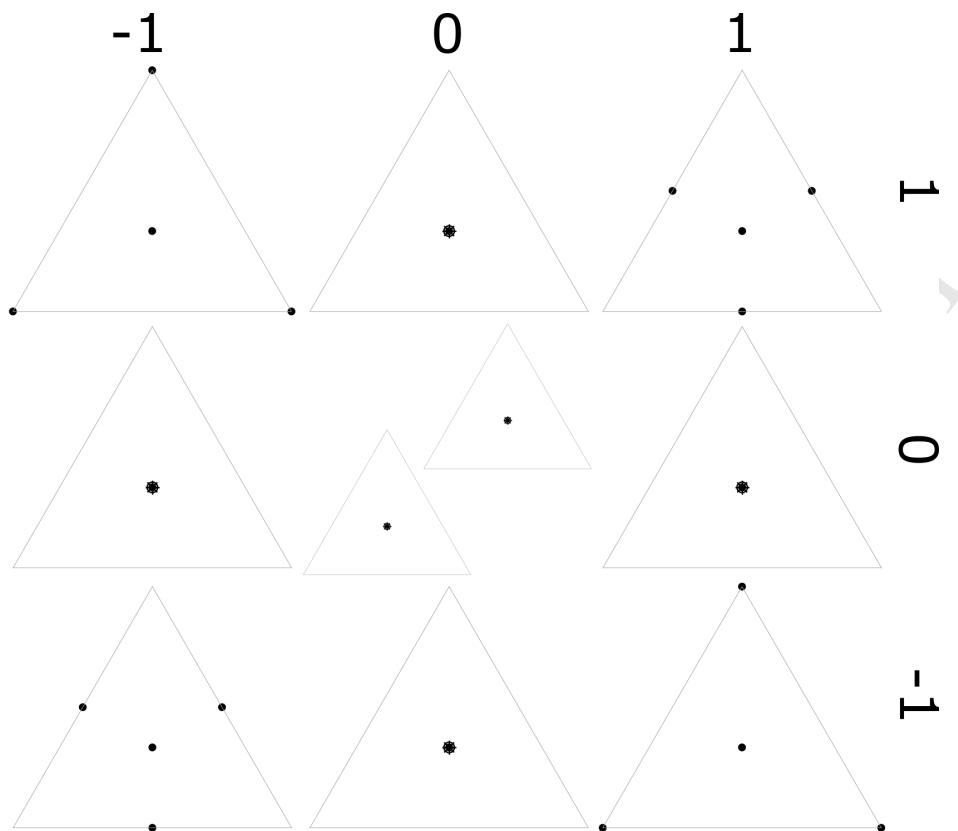


Figure 7: Benchmark split-plot design.

Combined markers indicate four replicated runs.

this kind of grouping occurs because many mixtures are incubated in the same climate chamber. For our proof-of-concept split-plot example, we use the 40-run split-plot design from Kowalski et al. (2002) as benchmark design. The 40 runs are evenly distributed over ten groups. The benchmark design and the D- and I-optimal split-plot designs are shown in Figures 7, 8 and 9. The benchmark split-plot design in Figure 7 involves the same kinds of design points as the benchmark completely randomized design in Figure 3, but it has many more replicates of the centroid of the experimental region for the mixture ingredients.

The model-driven nature of the D-optimal design is again clear when looking at Figure 8. Since there are only binary blending terms in the model, no ternary blends are included in the design. One binary blend deviates from the edge centroid when using the D-optimal design. Additionally, two of the nine process-variable combinations of the 3^2 factorial de-

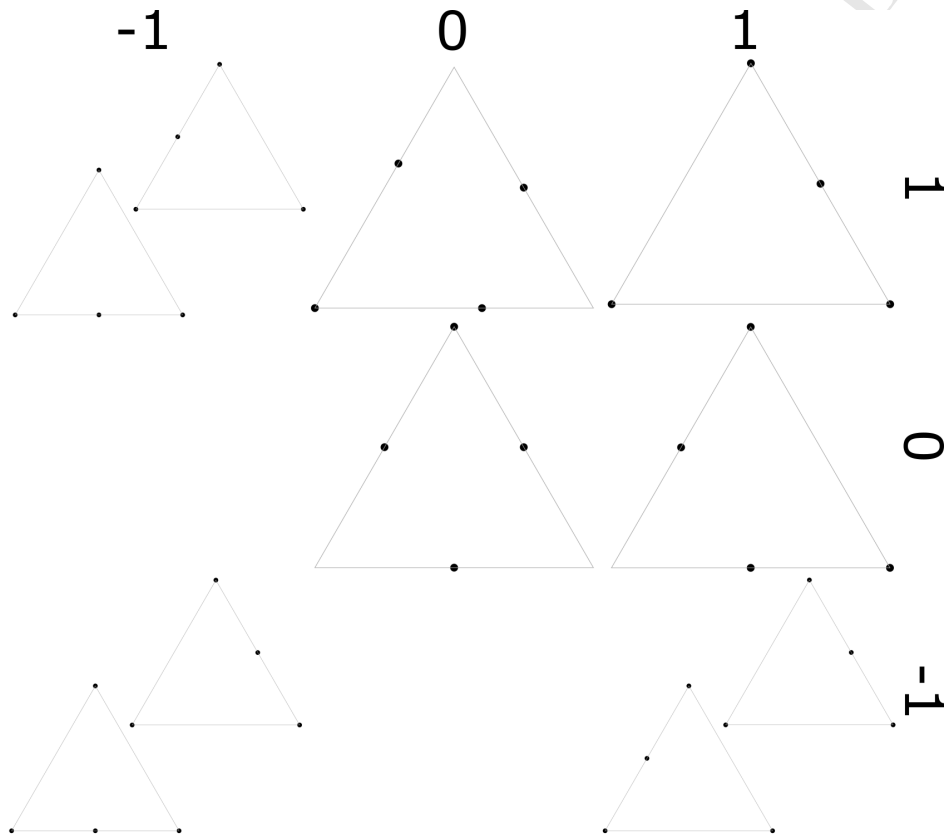


Figure 8: D-optimal split-plot design.

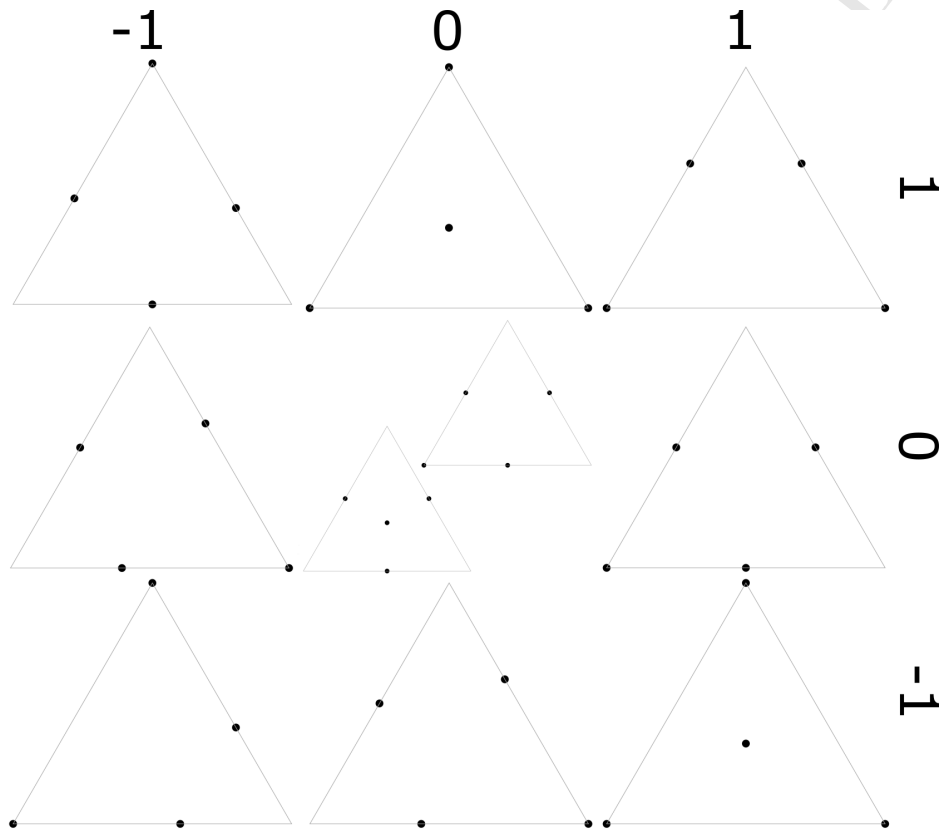


Figure 9: I-optimal split-plot design.

Table 2: Diagnostics and relative efficiencies of the three split-plot designs under comparison.

	$ \Sigma $	Max var	Avg var		Rel D-eff	Rel I-eff
Benchmark	2.04×10^{-1}	3.1202	0.7139	I-optimal to D-optimal	0.87	1.50
D-optimal	4.35×10^{-6}	1.6765	0.8102	Benchmark to D-optimal	0.49	1.14
I-optimal	3.25×10^{-5}	1.4938	0.5414	Benchmark to I-optimal	0.56	0.76

sign are not explored. Three of the process-variable combinations, involving the extreme levels for the process variables, are visited two times.

Unlike the D-optimal design, the I-optimal design shown in Figure 9 explores all combinations of the 3^2 factorial design for the process variables as well as the center of the mixture region. So, the I-optimal design does involve a ternary blend. Unlike the benchmark design, the I-optimal design only uses three runs to explore the center of the mixture region. These three runs appear at different settings of the process variables. There are three replicated runs, all of which appear at the $p_1 = p_2 = 0$ setting of the two process variables. In the I-optimal design, nine binary blends deviate from the edge centroids.

Table 2 reports the relative efficiencies of the three designs, assuming that the variance components corresponding to the individual runs and the groups both equal one. Both the benchmark and the I-optimal design have a smaller average variance of prediction than the D-optimal design. The I-optimal design has the smallest maximum variance of prediction of all three designs. In terms of D-optimality, the I-optimal design performs well relative to the D-optimal one. The benchmark design, however, performs poorly in terms of D-optimality, relative to both the D-optimal and the I-optimal design. In terms of I-optimality, the benchmark design is inferior to the I-optimal design, but it outperforms the D-optimal design.

The Fraction of Design Space plot in Figure 10 supports the findings in Table 2. The I-optimal split-plot design has the smallest variance of prediction over almost the entire experimental region. The benchmark design's variance of prediction is the least stable, so that that design has the largest spread in prediction variance. For a small portion of the

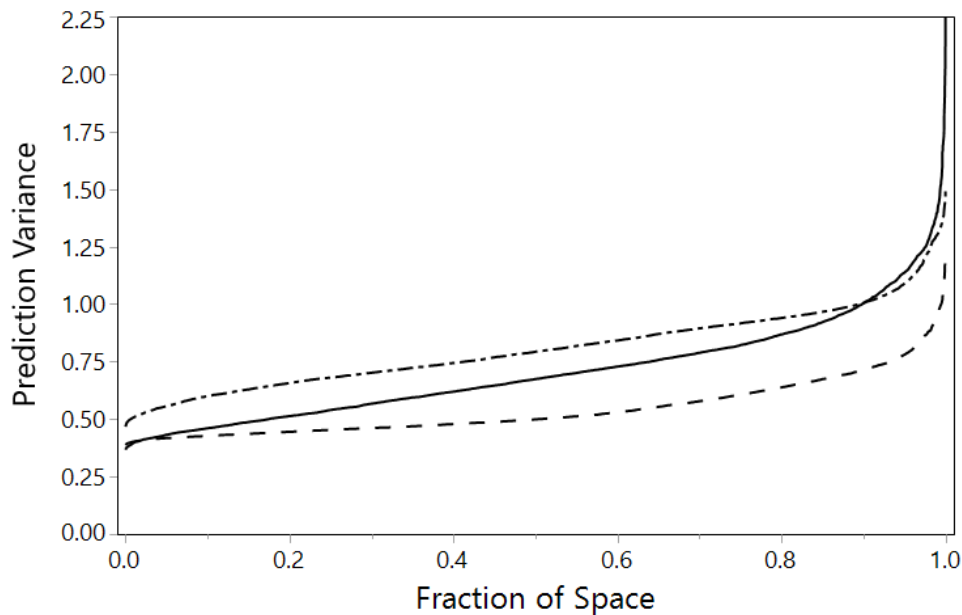


Figure 10: Fraction of Design Space plot for the benchmark (solid), D-optimal (dash-dotted) and I-optimal (dashed) split-plot design.

experimental region, the benchmark design has a larger variance of prediction than the D-optimal design. Comparing Figure 10 with Figure 6 shows that the difference between the benchmark design and the I-optimal design is larger in the split-plot scenario than in the completely randomized scenario.

When choosing a split-plot mixture-process variable design, this proof-of-concept example shows that, as for the completely randomized experiment, an I-optimal design represents the best trade-off between estimation precision, on the one hand, and a small variance of prediction over the entire experimental region, on the other hand. As for the completely randomized example, the *a priori* selected model can be extended by adding various higher-order terms when selecting the D- or I-optimal design. This is impossible when using the benchmark design.

6.3 Strip-plot designs

It is possible that, in addition to the grouping due to the process variables, there is grouping due to the mixtures. This happens when mixtures are produced in batches and each

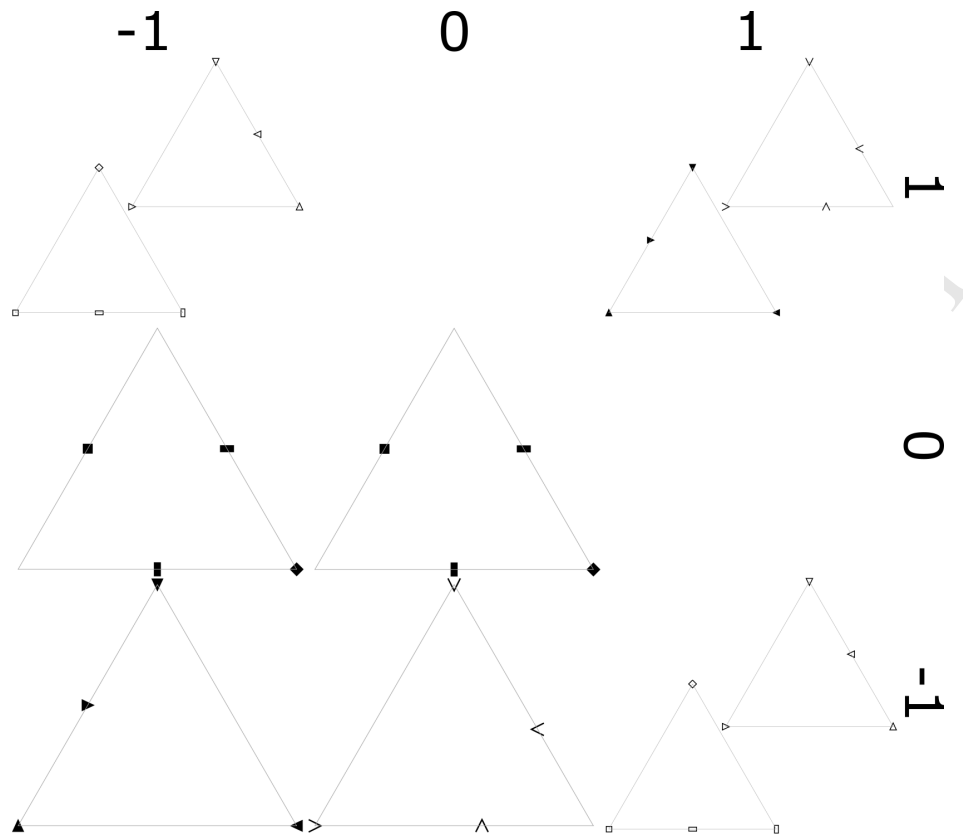


Figure 11: D-optimal strip-plot design.

Markers indicate the 20 batches.

batch is used for multiple observations. The resulting experiment then involves a strip-plot design. As there are no benchmark strip-plot designs in the literature, we only compare a D-optimal and an I-optimal strip-plot design here. The two designs involve 40 runs, 10 independent settings for the levels of the process variables, and 20 batches of a certain mixture, each of which is used for two runs. Consequently, the runs are grouped in two ways. There are 10 groups of the first kind, and 20 groups of the second kind.

The D-optimal design shown in Figure 11 has a few important characteristics. Firstly, just like in the D-optimal split-plot design, only seven of the nine process-variable combinations of the 3^2 factorial design are explored, and three process-variable combinations are replicated. Second, the center of the mixture region is not explored. This is again due to the model-driven nature of the D-optimal design, and leads to a large variance of prediction

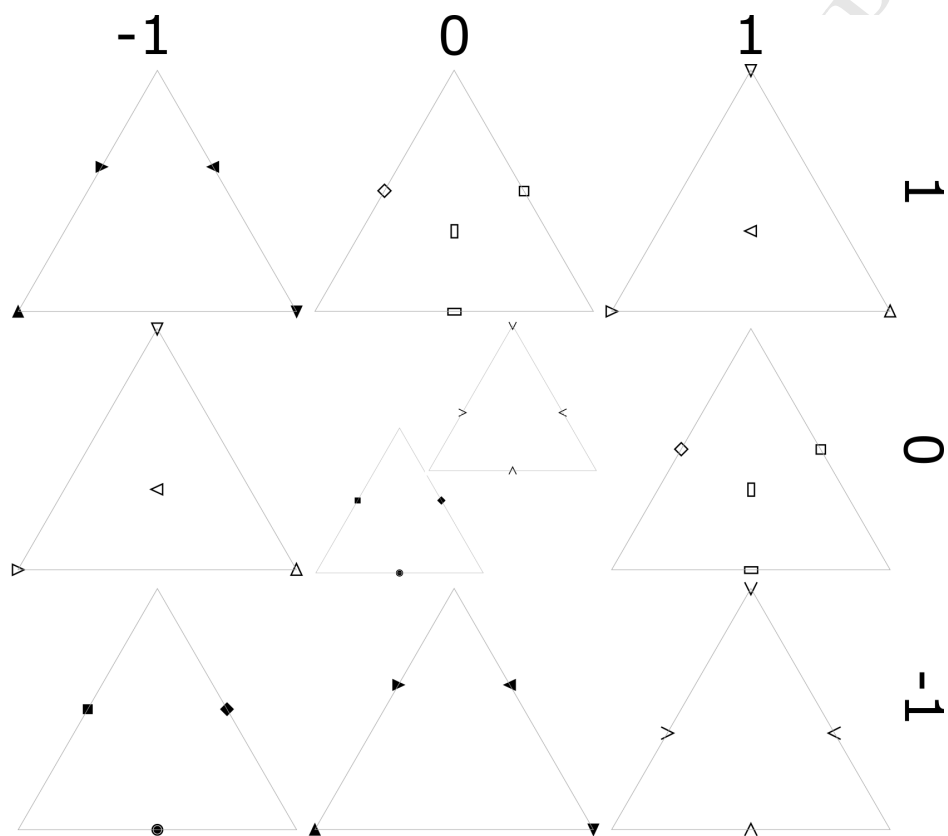


Figure 12: I-optimal strip-plot design.

Markers indicate the 20 batches.

The bullet in the bottom-left simplex is a replicated design point.

Table 3: Diagnostics and relative efficiencies of the three strip-plot designs under comparison.

	$ \Sigma $	Max var	Avg var		Rel D-eff	Rel I-eff
D-optimal	3.62×10^{-3}	2.4658	1.1417	I-optimal to D-optimal	0.79	1.51
I-optimal	1.36×10^{-1}	2.6247	0.7565	D-optimal to I-optimal	1.27	0.66

in the center of the mixture region. Third, several binary formulations deviate from the edge centroids. Finally, we see that each set of four formulations is used at two different combinations of process-variable levels.

The I-optimal design shown in Figure 12 does explore all combinations of the 3^2 factorial design for the process variables. In addition, it uses two batches to explore the mixture region's centroid to decrease the variance of prediction in the center of the mixture region. In the I-optimal design, four binary mixtures are shifted away from the edge centroid. As for the D-optimal design, each set of four formulations is used at two different combinations of levels of the process variables.

Table 3 compares the diagnostics and relative efficiencies of the two strip-plot designs. The I-optimal design clearly has the smallest average variance of prediction. In terms of the the maximum variance of prediction, the D-optimal and I-optimal designs hardly differ. In terms of D-optimality, the I-optimal design performs relatively well compared to the D-optimal design. In terms of I-optimality, the I-optimal design substantially outperforms the D-optimal design. The Fraction of Design Space plot in Figure 13 supports the findings in Table 3. The I-optimal design has the smallest variance of prediction over almost the entire design region.

If a strip-plot mixture-process variable experiment is conducted, our proof-of-concept example shows that an I-optimal design is the most sensible choice, as for the two other scenarios discussed. The I-optimal design combines a small variance of prediction over almost the entire experimental region with a reasonably good estimation precision. Both models can be extended with cross-products of the process variables and the non-linear blending terms. In addition, using the I-optimal design the model can be extended with a

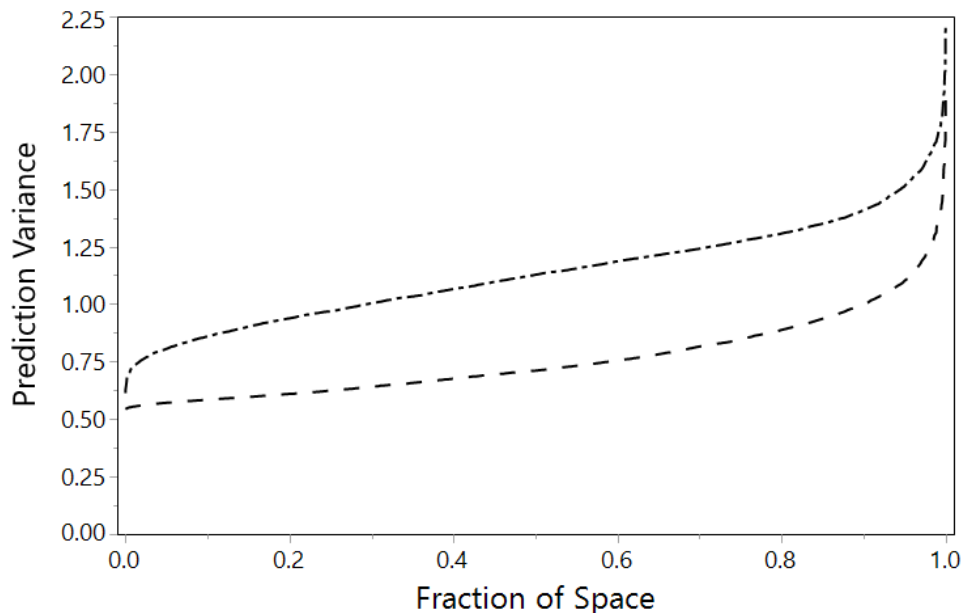


Figure 13: Fraction of Design Space plot for the D-optimal (dash-dotted) and I-optimal (dashed) strip-plot designs.

ternary blending term. This is impossible when selecting a D-optimal design.

7 Excipient compatibility experiment

7.1 Model and design

As the *a priori* model for constructing a design for the excipient compatibility study described in Section 2, we used the mixture-process variable model in Equation (7). There were seven mixture components ($q = 7$) and two process variables ($r = 2$) in the experiment. The storage conditions were imposed on groups of 15 formulations and all formulations were produced independently using a robot, so that a split-plot design was required. Only six climate chambers were available, so that the total number of experimental tests was 90 and the number of groups defined by the settings of the process variables was six. Importantly, six independent settings of the process variables are insufficient to estimate a complete two-factor response-surface model with two main effects, one two-factor interaction and two quadratic effects. Kowalski et al. (2002) discuss a similar scenario, and suggest dropping the quadratic terms in the process variables from the response surface model. Leaving

out the quadratic terms may not be warranted in certain applications and should therefore be avoided whenever possible. For the excipient compatibility experiment discussed here, however, the subject-matter specialists did not expect curvature, making it sensible to drop these terms. Eventually, the following 43-term mixture-process variable model was used for the design optimization:

$$E(Y) = \sum_{i=1}^7 \beta_i m_i + \sum_{i=1}^6 \sum_{j=i+1}^7 \beta_{ij} m_i m_j + \sum_{i=1}^7 \sum_{k=1}^2 \gamma_{ik} m_i p_k + \alpha_1 p_1 p_2. \quad (15)$$

The goals of an excipient compatibility study are to minimize the degradation of the API and to make the degradation as robust as possible to different storage conditions. This suggests that the emphasis is on predictive modeling and optimization rather than on a precise model estimation. In addition, it was unclear how to adapt the designs proposed by Kowalski et al. (2002) to cope with the constrained experimental region defined in Equations (1)–(3). Following our recommendations in Section 6, we constructed an I-optimal split-plot design using the coordinate-exchange algorithm in JMP 12. The final design is shown in Table 4. A major difference between the design for the motivating example and those for the proof-of-concept examples is that the former design only involves two levels for the process variables relative humidity and temperature, while the latter designs involved three levels for each of these variables. This is due to the fact that we did not incorporate quadratic effects for the process variables in the model in Equation (15).

Because the response had a right-skewed distribution and considerable heteroscedasticity was found in the residuals, we applied a logarithmic transformation to the responses. This was sufficient to remove most of the heteroscedasticity from the residuals.

7.2 Data analysis

We estimated the parameters from the model in Equation (15) using generalized least squares. For the two variance components, describing the group-to-group variation and the run-to-run variation, we used REML. To find a parsimonious model for the responses with a good fit, we performed backward selection based on the corrected Akaike information

Table 4: 90-run I-optimal design for the excipient compatibility study.

Group	m_1	m_2	m_3	m_4	m_5	m_6	m_7	$p_1 (T)$	$p_2 (RH)$	$\log(Y)$
1	0.00	0.59	0.00	0.30	0.00	0.00	0.11	1	1	-0.10536
1	0.00	0.00	0.81	0.19	0.00	0.00	0.00	1	1	-0.40048
1	0.11	0.00	0.30	0.59	0.00	0.00	0.00	1	1	-0.57982
1	0.22	0.33	0.00	0.00	0.22	0.07	0.15	1	1	-0.44629
1	0.85	0.00	0.00	0.00	0.00	0.00	0.15	1	1	-0.79851
1	0.00	0.70	0.00	0.00	0.00	0.00	0.30	1	1	-0.86750
1	0.00	0.70	0.00	0.00	0.00	0.00	0.30	1	1	-0.86750
1	0.11	0.11	0.15	0.00	0.26	0.37	0.00	1	1	-1.23787
1	0.41	0.41	0.00	0.00	0.00	0.19	0.00	1	1	-1.04982
1	0.15	0.15	0.15	0.00	0.56	0.00	0.00	1	1	-0.84397
1	0.11	0.11	0.15	0.26	0.26	0.00	0.11	1	1	-0.27444
1	0.30	0.00	0.33	0.07	0.00	0.15	0.15	1	1	-0.27444
1	0.30	0.15	0.15	0.00	0.00	0.11	0.30	1	1	-0.38566
1	0.00	0.41	0.00	0.22	0.00	0.37	0.00	1	1	0.41211
1	0.44	0.11	0.15	0.07	0.22	0.00	0.00	1	1	-0.51083
2	0.00	0.30	0.33	0.19	0.00	0.07	0.11	-1	-1	-0.65393
2	0.78	0.00	0.00	0.00	0.22	0.00	0.00	-1	-1	-1.51413
2	0.48	0.37	0.00	0.00	0.00	0.00	0.15	-1	-1	-0.91629
2	0.00	0.37	0.22	0.15	0.00	0.11	0.15	-1	-1	-0.63488
2	0.00	0.00	0.56	0.26	0.19	0.00	0.00	-1	-1	-0.75502
2	0.07	0.26	0.26	0.00	0.15	0.15	0.11	-1	-1	-0.91629
2	0.70	0.00	0.15	0.00	0.00	0.15	0.00	-1	-1	-1.02165
2	0.33	0.30	0.00	0.37	0.00	0.00	0.00	-1	-1	-0.91629
2	0.41	0.00	0.00	0.00	0.00	0.30	0.30	-1	-1	-1.30933
2	0.22	0.07	0.33	0.00	0.26	0.00	0.11	-1	-1	-0.94161
2	0.41	0.00	0.00	0.00	0.59	0.00	0.00	-1	-1	-0.99425
2	0.74	0.00	0.00	0.26	0.00	0.00	0.00	-1	-1	-0.89160
2	0.41	0.00	0.00	0.41	0.00	0.00	0.19	-1	-1	-0.86750
2	0.41	0.00	0.00	0.33	0.00	0.26	0.00	-1	-1	-0.79851
2	0.41	0.00	0.00	0.00	0.26	0.33	0.00	-1	-1	-1.46968
3	0.41	0.00	0.00	0.00	0.30	0.00	0.30	-1	-1	-0.99425
3	0.00	0.30	0.11	0.41	0.00	0.19	0.00	-1	-1	-0.35667
3	0.00	0.00	0.48	0.00	0.00	0.37	0.15	-1	-1	-1.17118
3	0.00	0.37	0.48	0.00	0.00	0.15	0.00	-1	-1	-1.04982
3	0.00	0.00	0.70	0.00	0.00	0.00	0.30	-1	-1	-1.27297
3	0.00	0.48	0.00	0.00	0.15	0.37	0.00	-1	-1	-0.96758
3	0.00	0.85	0.00	0.15	0.00	0.00	0.00	-1	-1	-1.38629
3	0.37	0.22	0.00	0.00	0.19	0.15	0.07	-1	-1	-1.46968
3	0.00	0.41	0.44	0.00	0.15	0.00	0.00	-1	-1	-1.30933
3	0.00	0.22	0.37	0.30	0.00	0.00	0.11	-1	-1	-1.42712
3	0.00	0.41	0.00	0.00	0.41	0.00	0.19	-1	-1	-1.30933
3	0.00	0.00	0.41	0.30	0.30	0.00	0.00	-1	-1	-1.13943
3	0.33	0.07	0.22	0.15	0.00	0.11	0.11	-1	-1	-0.96758
3	0.00	0.00	0.41	0.00	0.48	0.11	0.00	-1	-1	-1.34707
3	0.26	0.00	0.59	0.15	0.00	0.00	0.00	-1	-1	-1.04982
4	0.30	0.00	0.44	0.00	0.26	0.00	0.00	1	1	-0.73397
4	0.26	0.00	0.15	0.30	0.00	0.30	0.00	1	1	0.19062
4	0.63	0.00	0.00	0.00	0.00	0.37	0.00	1	1	-0.77653
4	0.41	0.00	0.00	0.22	0.30	0.07	0.00	1	1	-0.16252
4	0.00	0.30	0.22	0.19	0.19	0.11	0.00	1	1	-0.06188
4	0.07	0.26	0.30	0.26	0.00	0.11	0.00	1	1	-0.02020
4	0.33	0.07	0.00	0.30	0.00	0.00	0.30	1	1	-0.75502
4	0.00	0.33	0.52	0.00	0.00	0.00	0.15	1	1	-1.04982
4	0.41	0.22	0.22	0.15	0.00	0.00	0.00	1	1	-0.63488
4	0.26	0.00	0.41	0.00	0.00	0.19	0.15	1	1	-0.24846
4	0.07	0.63	0.00	0.00	0.30	0.00	0.00	1	1	-0.91629
4	0.00	0.19	0.22	0.00	0.00	0.30	0.30	1	1	-0.99425
4	0.22	0.19	0.00	0.26	0.00	0.19	0.15	1	1	-0.13926
4	0.00	0.59	0.19	0.00	0.00	0.22	0.00	1	1	-1.13943
4	0.00	0.00	0.41	0.00	0.37	0.00	0.22	1	1	0.02956
5	0.33	0.00	0.19	0.15	0.26	0.00	0.07	-1	1	-0.59784
5	0.00	0.26	0.15	0.00	0.41	0.19	0.00	-1	1	-1.23787
5	0.00	0.00	0.41	0.00	0.22	0.19	0.19	-1	1	-1.23787
5	0.11	0.00	0.44	0.15	0.00	0.30	0.00	-1	1	-0.67334
5	0.07	0.67	0.11	0.00	0.15	0.00	0.00	-1	1	-1.30933
5	0.33	0.00	0.07	0.19	0.19	0.15	0.07	-1	1	-1.34707
5	0.30	0.26	0.07	0.07	0.07	0.11	0.11	-1	1	-1.27297
5	0.00	0.48	0.00	0.00	0.00	0.37	0.15	-1	1	-1.46968
5	0.41	0.30	0.00	0.15	0.00	0.15	0.00	-1	1	-1.27297
5	0.63	0.00	0.00	0.00	0.07	0.00	0.30	-1	1	-1.23787
5	0.30	0.00	0.56	0.00	0.00	0.00	0.15	-1	1	-1.42712
5	0.00	0.37	0.07	0.00	0.15	0.11	0.30	-1	1	-1.20397
5	0.00	0.00	0.41	0.30	0.00	0.00	0.30	-1	1	-1.20397
5	0.00	0.00	0.67	0.00	0.15	0.19	0.00	-1	1	-1.60944
5	0.00	0.37	0.00	0.48	0.15	0.00	0.00	-1	1	-1.30933
6	0.19	0.19	0.19	0.07	0.00	0.37	0.00	1	-1	0.31481
6	0.00	0.41	0.00	0.59	0.00	0.00	0.00	1	-1	-0.57982
6	0.11	0.00	0.70	0.00	0.07	0.11	0.00	1	-1	-0.91629
6	0.00	0.37	0.00	0.00	0.15	0.30	0.19	1	-1	-0.44629
6	0.00	0.41	0.00	0.30	0.00	0.00	0.30	1	-1	-0.82098
6	0.00	0.00	0.41	0.30	0.00	0.19	0.11	1	-1	-1.07881
6	0.44	0.07	0.00	0.15	0.19	0.15	0.00	1	-1	-0.99425
6	0.00	0.81	0.00	0.00	0.00	0.19	0.00	1	-1	-1.23787
6	0.33	0.41	0.00	0.11	0.00	0.00	0.15	1	-1	-0.82098
6	0.00	0.22	0.33	0.00	0.15	0.00	0.30	1	-1	-1.02165
6	0.00	0.26	0.15	0.00	0.48	0.11	0.00	1	-1	-1.27297
6	0.37	0.00	0.11	0.00	0.26	0.11	0.15	1	-1	-0.12783
6	0.00	0.41	0.00	0.30	0.30	0.00	0.00	1	-1	-0.44629
6	0.48	0.00	0.19	0.15	0.00	0.00	0.19	1	-1	-0.61619
6	0.00	0.41	0.33	0.00	0.26	0.00	0.00	1	-1	-0.91629

criterion (AICc). We used the AICc as a criterion for the model selection rather than the p-value for two reasons. First, we designed the experiment with predictive modelling in mind, and inferences on individual parameters are more difficult in a mixture context. Second, the AICc provides a tradeoff between underfitting, as it seeks a model that minimizes the SSE, and overfitting, by penalizing for the number of terms in the model. During the model selection, it is of utmost importance that all linear blending terms remain in the model and that the main effects of the process variables defining the storage conditions are excluded as long as their cross-products with the component proportions are in the model.

The first model used to describe the system was the mixture-process variables model given in Equation (15). However, backward selection resulted in a non-satisfactory model with a small R_a^2 and relatively high $RMSE$. Therefore, starting from the initial model, we performed a forward selection based on the AICc and included interactions between the process variables and the non-linear blending terms, as in Equation (8). If all of these terms would be included, the full model would involve 85 instead of 43 terms:

$$E(Y) = \sum_{i=1}^7 \beta_i m_i + \sum_{i=1}^6 \sum_{j=i+1}^7 \beta_{ij} m_i m_j + \sum_{i=1}^7 \sum_{k=1}^2 \gamma_{ik} m_i p_k + \sum_{i=1}^6 \sum_{j=i+1}^7 \sum_{k=1}^2 \gamma_{ijk} m_i m_j p_k + \alpha_{12} p_1 p_2. \quad (16)$$

Table 5 reports some goodness-of-fit statistics for the full model given in (16) and the reduced model selected based on the AICc. The coefficient of determination (R^2) quantifies the proportion of variation in the response (i.e., the log degradation) that can be explained by the model, and the adjusted coefficient of determination (R_a^2) corrects the R^2 for the number of parameters in the model. The AICc value measures the information lost when using a certain model to represent the data. For the AICc criterion, small values are desirable, while large values are desirable for the R^2 and R_a^2 values. A common feature of the R_a^2 and AICc criteria is that they prefer parsimonious models (i.e., models with fewer terms), all other things being equal. The variance ratio is the ratio of the group-to-group variance (σ_λ^2) and the run-to-run variance (σ_ϵ^2). If this ratio exceeds one, most of the random variability is between the different groups in the experiment. Otherwise, most of the

Table 5: Goodness-of-fit statistics for the fitted models for the excipient compatibility study.

	Full	Reduced
No. of terms	84	31
R^2	0.988	0.877
R_a^2	0.821	0.815
AICc	4886.797	41.288
Variance ratio	0.826	0.370

variability is between the individual runs. When the variance ratio is close to zero, this indicates that the correlation between responses obtained within the same group is small.

Two major conclusions can be drawn from Table 5. First, almost two thirds of the terms in the full mixture-process variable model can be dropped. The resulting reduced model has a fit similar to that of the full model, in terms of the R^2 and R_a^2 value. Based on the AICc value, the reduced model is by far the better option of the two. Second, for both models, the variance ratio is smaller than one. In other words, the estimated variance in the responses due to the groups is smaller than the estimated run-to-run variance. This indicates that only a limited portion of the observed variation in the (log) degradation between the different groups or climate chambers cannot be explained by the levels of the process variables. So, for this particular data set, ignoring the grouping due to the climate chambers and treating the 90 runs are nearly independent would be a minor mistake. Nevertheless, taking into account the split-plot nature of the experiment in the data analysis is necessary, because it is not known in general how large the group-to-group variation is in excipient compatibility studies.

The estimates of the parameters in the reduced model are given in Table 6. The model involves many interaction effects between linear and non-linear blending terms, on the one hand, and the process variables, on the other hand, indicating that the effects of the mixture component proportions differ significantly from one storage condition to another, and that the effects of temperature and relative humidity differ significantly from one mixture to another. The fact that the significances of these interaction terms can be established is due to the fact that the excipient compatibility study involved an informative experimen-

tal design and a model intended to investigate the simultaneous effect of the component proportions and the process variables.

When analyzing data from mixture experiments, it is customary to use contour plots for interpreting the effects. Contour plots corresponding to the reduced model appear in Figures 14 and 15. The plots at the bottom/top of the figures correspond to the low/high level of the first process variable p_1 (Temperature), while those at the left/right correspond to the low/high level of the second process variable p_2 (Relative Humidity). Because the proportion of magnesium stearate (m_7) is expected to be small in the final formulation, we fix it at 0.08 and exclude it from our further analysis. The given value is the expected degradation of the API, expressed in %.

Figure 14 shows the effects of the three fillers when the excipient proportions m_4 , m_5 , m_6 and m_7 are all set to 0.08. This proportion corresponds to the levels of m_4 , m_5 , m_6 and m_7 near the centroid of the experimental region. Figure 15 shows the effect of three of the four functional excipients, m_4 , m_5 and m_6 , in the event $m_1 = m_2 = m_3 = 0.227$ and $m_7 = 0.08$. The filler proportion of 0.227 is chosen such that the range for the three functional excipients, other than magnesium stearate, is maximal.

When looking at Figure 14 we see that for low temperatures, using a larger proportion of the first filler (m_1) leads to the least degradation. Remarkably, at low temperature, a higher humidity results in less degradation. At low humidity levels, there is an active synergistic effect between the three fillers causing more degradation in the center of the mixture region. For high temperature and low humidity, using a larger proportion of the third filler (m_3) results in the least degradation. For this storage condition there is little difference in degradation between the two first fillers. At high humidity, there is little difference between the three fillers. For both humidity levels, at high temperature there is a strong synergistic effect between the first and the third filler. So, any combination of these two fillers should be avoided.

Table 6: Parameter estimates, standard errors, t -ratios and p -values for the reduced model.

Term	Estimate	Std Error	t -Ratio	Prob $> t $
m_1	- 1.5312	0.1985	.	.
m_2	- 1.2277	0.1332	.	.
m_3	- 1.3460	0.1717	.	.
m_4	- 3.4294	0.6127	.	.
m_5	- 0.7845	0.2320	.	.
m_6	- 0.5155	0.2952	.	.
m_7	- 0.8496	0.3849	.	.
$m_1 \times m_3$	2.8085	0.7789	3.6058	0.0007
$m_1 \times m_4$	4.2376	1.4126	2.9998	0.0040
$m_2 \times m_4$	6.3895	1.3842	4.6162	0.0000
$m_3 \times m_4$	5.3522	1.3749	3.8928	0.0003
$m_4 \times m_5$	5.0709	1.4951	3.3917	0.0013
$m_1 \times m_6$	- 0.9762	0.8980	-1.0870	0.2817
$m_4 \times m_6$	11.0919	1.6112	6.8841	0.0000
$m_5 \times m_6$	- 3.4308	1.6938	-2.0256	0.0475
$m_1 \times m_7$	2.9613	1.1338	2.6118	0.0115
$m_3 \times m_7$	- 0.7260	1.1106	-0.6537	0.5159
$m_5 \times m_7$	5.5053	1.6613	3.3138	0.0016
$m_1 \times p_1$	- 0.1938	0.1882	-1.0300	0.3072
$m_6 \times p_1$	0.1920	0.2405	0.7983	0.4286
$m_1 \times m_3 \times p_1$	1.6450	0.6782	2.4257	0.0184
$m_2 \times m_4 \times p_1$	2.4481	0.5195	4.7129	0.0000
$m_1 \times m_6 \times p_1$	4.9865	1.2976	3.8429	0.0003
$m_1 \times m_7 \times p_1$	2.3514	0.9553	2.4614	0.0169
$m_5 \times m_7 \times p_1$	5.7172	1.3487	4.2390	0.0001
$m_1 \times m_3 \times p_2$	- 1.0953	0.6647	-1.6477	0.1048
$m_3 \times m_4 \times p_2$	2.4720	0.6369	3.8813	0.0003
$m_1 \times m_5 \times p_2$	1.4603	0.6756	2.1613	0.0348
$m_1 \times m_6 \times p_2$	- 2.5250	0.9504	-2.6569	0.0102
$m_2 \times m_6 \times p_2$	- 1.5166	0.5959	-2.5450	0.0136
$p_1 \times p_2$	0.1204	0.0573	2.1002	0.1180

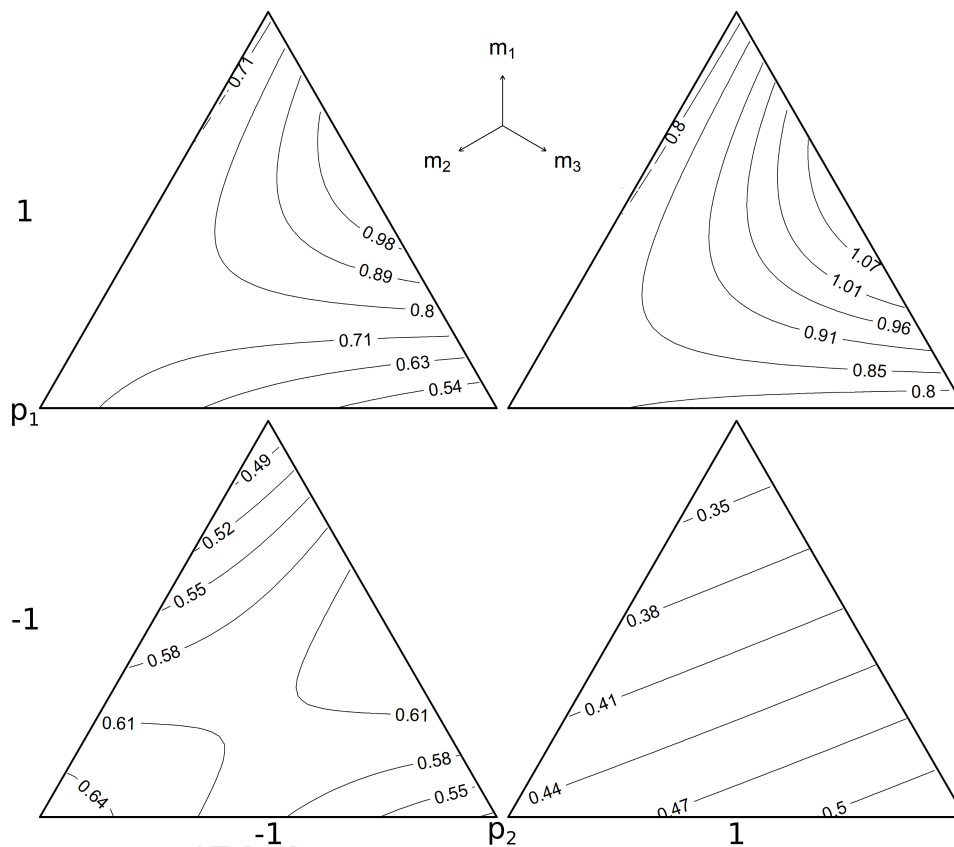


Figure 14: Contour plots for the fillers m_1 , m_2 , and m_3 when $m_4 = m_5 = m_6 = m_7 = 0.08$.

The top (bottom) plots are for a large (small) relative humidity.

The left (right) plots are for a low (high) temperature.

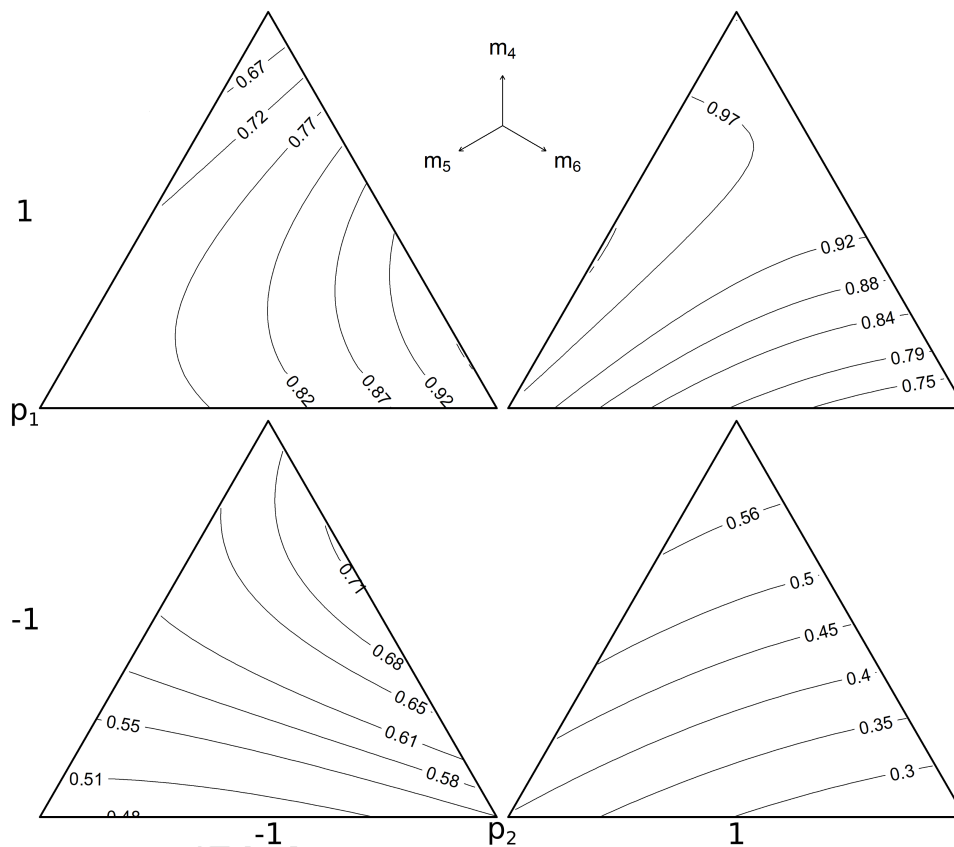


Figure 15: Contour plots for the functional exponents m_4 , m_5 , and m_6 when $m_1 = m_2 = m_3 = 0.227$ and $m_7 = 0.08$. The top (bottom) plots are for a large (small) relative humidity. The left (right) plots are for a low (high) temperature.

Figure 15 shows that the interactions between the functional excipients and the storage conditions are more pronounced. In general, a higher temperature results in larger degradations. At low temperature and low humidity, a larger proportion of the desintegrant (m_5) results in smallest degradation, while a non-zero proportion of the glidant (m_4) generally results in larger degradations. The active synergistic effect between the proportion of the glidant (m_4) and the surfactant (m_6) causes a maximum degradation near the vertex of the glidant. However, at high temperature and low humidity setting, high glidant proportions (m_4) leads to a small degradation and a non-zero proportion of the surfactant (m_6) results in larger degradations. The largest degradation is now found close to the vertex of the surfactant (m_6). At high humidity, the degradation is minimal when $m_6 = 1$ at low and high temperatures. The degradation is always maximal when the glidant has a non-zero proportion. At high temperature there is an active synergistic effect between the glidant (m_4) and the desintegrant (m_5), so that a combination of these excipients result in larger degradations.

8 Conclusion

We have shown that mixture-process variable models and experimental designs are useful tools for excipient compatibility studies. Compared to the traditional excipient compatibility study approach, these models and designs allow the incorporation of the storage conditions, temperature and relative humidity. A major benefit is that this methodology allows us to test whether the impact of the different excipients on the degradation depends on the temperature and the humidity, and whether the impact of temperature and humidity depends on the excipient proportions used in the formulation. In the statistical jargon, the methodology allows us to test whether there are interaction effects between excipient proportions, on the one hand, and storage conditions, on the other hand.

One challenge with the mixture-process variable methodology is to produce a suitable experimental design. In this article, we studied three different approaches to design excipient compatibility studies, using three different scenarios: a completely randomized experiment, a split-plot experiment and a strip-plot experiment. We compare designs proposed

by Kowalski et al. (2000, 2002) to D- and I-optimal designs. The optimal design approach allows us to take into account any constraints on the component proportions, both linear and nonlinear. This flexibility was of utmost importance in our motivating example. Another advantage is that optimal designs can be computed for any kind of grouping in the experimental setup. So, it is possible to compute optimal split-plot and strip-plot designs, in addition to optimal completely randomized designs. It turned out that the I-optimal designs provide an excellent trade-off between estimation precision and predictive capability. Moreover, for the examples discussed here, the I-optimal design allows additional interactions between the process variables and the nonlinear blending terms to be included in the model, even though these terms were not included in the *a priori* model used for the design construction. Including these terms increased the goodness of fit of our model for the motivating example substantially.

We demonstrated the I-optimal split-plot design approach using an example of an excipient compatibility study with seven excipients. By means of contour plots, we found many interesting active synergistic effects and it became possible to visualize the effect of the storage conditions. The larger degradations at high temperatures was expected, but the lower degradations at low temperature and high humidity was not. This has to be investigated more into detail. For further formulation development, no real incompatibilities are found, but combining the fillers mannitol (m_1) and lactose (m_3) should be avoided.

Acknowledgement

The research described in this article was funded by Janssen Pharmaceutica.

References

Anderson-Cook, C. M., Goldfarb, H. B., Borror, C. M., Montgomery, D. C., Canter, K. G. and Twist, J. N. (2004). Mixture and mixture-process variable experiments for pharmaceutical applications, *Pharmaceutical Statistics* **3**(4): 247–260.

- Arnouts, H. and Goos, P. (2015). Staggered-level designs for response surface modeling, *Journal of Quality Technology* **47**: 156–175.
- Arnouts, H., Goos, P. and Jones, B. (2010). Design and analysis of industrial strip-plot experiments, *Quality and Reliability Engineering International* **26**: 127–136.
- Arnouts, H., Goos, P. and Jones, B. (2013). Three-stage strip-plot designs, *Journal of Quality Technology* **45**: 1–17.
- Aulton, M. E. (2013). *Aulton's pharmaceuticals: the design and manufacture of medicines*, 4th edn, Churchill Livingstone, 2013, Edinburgh.
- Bisgaard, S. (2000). The design and analysis of $2^{(k-p)} \times 2^{(q-r)}$ split-plot experiments, *Journal of Quality Technology* **32**(1): 39–56.
- Chipman, H., Hamada, M. and Wu, C. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing, *Technometrics* **39**(4): 372–381.
- Cornell, J. A. (1988). Analyzing data from mixture experiments containing process variables: a split-plot approach, *Journal of Quality Technology* **20**(1): 2–23.
- Cornell, J. A. (2002). *Experiments with mixtures : designs, models, and the analysis of mixture data*, 3th edn, Wiley, New York.
- Cornell, J. A. and Gorman, J. W. (1984). Fractional design plans for process variables in mixture experiments, *Journal of Quality Technology* **16**(1): 20–38.
- Draper, N. R. and Lawrence, W. E. (1965a). Mixture designs for four factors, *Journal of the Royal Statistical Society. Series B (Methodological)* **27**(3): 473–478.
- Draper, N. R. and Lawrence, W. E. (1965b). Mixture designs for three factors, *Journal of the Royal Statistical Society. Series B (Methodological)* **27**(3): 450–465.
- Goos, P. and Donev, A. N. (2007). Tailor-made split-plot designs for mixture and process variables, *Journal of Quality Technology* **39**(4): 326–339.
- Goos, P. and Jones, B. (2011). *Optimal design of experiments: a case study approach*, Wiley, 2011, Chichester.

- Goos, P., Jones, B. and Syafitri, U. (2016). I-optimal design of mixture experiments, *Journal of the American Statistical Association* **111**(514): 899–911.
- Goos, P., Langhans, I. and Vandebroek, M. (2006). Practical inference from industrial split-plot designs, *Journal of Quality Technology* **38**(2): 162–179.
- Goos, P. and Syafitri, U. (2014). V-optimal mixture designs for the q th degree model, *Chemometrics and Intelligent Laboratory Systems* **136**: 173–178.
- International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (2003). Stability testing of new drug substances and products - Q1A (R2).
- Jones, B. and Goos, P. (2007). A candidate-set-free algorithm for generating D-optimal split-plot designs, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **56**(3): 347–364.
- Jones, B. and Goos, P. (2012). I-optimal versus D-optimal split-plot response-surface designs, *Journal of Quality Technology* **44**: 85–101.
- Jones, B. and Goos, P. (2015). Optimal design of blocked experiments in the presence of supplementary information about the blocks, *Journal of Quality Technology* **47**: 301–317.
- Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood, *Biometrics* **53**(3): 983–997.
- Kenward, M. G. and Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood, *Computational Statistics and Data Analysis* **53**(7).
- Kowalski, S. M., Cornell, J. A. and Vining, G. G. (2000). A new model and class of designs for mixture experiments with process variables, *Communications in Statistics - Theory and Methods* **29**(9-10): 2255–2280.
- Kowalski, S. M., Cornell, J. A. and Vining, G. G. (2002). Split-plot designs and estimation methods for mixture experiments with process variables, *Technometrics* **44**(1): 72–79.
- Langhans, I., Goos, P. and Vandebroek, M. (2005). Identifying effects under a split-plot design structure, *Journal of Chemometrics* **19**(1): 5–15.

- Letsinger, J. D., Myers, R. H. and Lentner, M. (1996). Response surface methods for bi-randomization structures, *Journal of Quality Technology* **28**(4): 381–397.
- Martinello, T., Kaneko, T. M., Velasco, M. V. R., Taqueda, M. E. S. and Consiglieri, V. O. (2006). Optimization of poorly compactable drug tablets manufactured by direct compression using the mixture experimental design, *International Journal of Pharmaceutics* **322**(1-2): 87–95.
- Montgomery, D. C. (2012). *Design and Analysis of Experiments, 8th Edition*, John Wiley & Sons.
- Næs, T., Færgestad, E. M. and Cornell, J. A. (1998). A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables, *Chemometrics and Intelligent Laboratory Systems* **41**(2): 221–235.
- Nyqvist, H. (1986). Preformulation studies of drug substances for solid dosage forms, *Drug Development and Industrial Pharmacy* **12**(7): 953–968.
- Piepel, G. F. (1988). Programs for generating extreme vertices and centroids of linearly constrained experimental regions, *Journal of Quality Technology* **20**(2): 125–139.
- Piepel, G. F. (1999). Modeling methods for mixture-of-mixtures experiments applied to a tablet formulation problem, *Pharmaceutical Development and Technology* **4**(4): 593–606.
- Piepel, G. F., Cooley, S. K. and Jones, B. (2005). Construction of a 21-component layered mixture experiment design using a new mixture coordinate-exchange algorithm, *Quality Engineering* **17**(4): 579–594.
- Rodríguez, M., Jones, B., Borrór, C. M. and Montgomery, D. C. (2010). Generating and assessing exact G-optimal designs, *Journal of Quality Technology* **42**(1): 3–20.
- Ruseckaite, A., Goos, P. and Fok, D. (2017). Bayesian D-optimal choice designs for mixtures, *Applied Statistics* **66**(2): 363–386.
- Scheffé, H. (1958). Experiments with mixtures, *Journal of the Royal Statistical Society. Series B (Methodological)* **20**(2): 344–360.
- Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures, *Journal of the Royal Statistical Society. Series B (Methodological)* **25**(2): 235–263.
- Shmueli, G. (2010). To explain or to predict?, *Statistical Science* **25**(3): 289–310.

Smith, W. F. (2005). *Experimental Design for Formulation*, ASA-SIAM Series on Statistics and Applied Probability, Society for Industrial and Applied Mathematics.

Syafitri, U., Sartono, B. and Goos, P. (2015). I-optimal design of mixture experiments in the presence of ingredient availability constraints, *Journal of Quality Technology* **47**: 220–234.

Waterman, K. C. and Adami, R. C. (2005). Accelerated aging: Prediction of chemical stability of pharmaceuticals, *International Journal of Pharmaceutics* **293**(1–2): 101–125.

Highlights**“Optimal design of experiments for excipient compatibility studies”**

- We discuss the typical split-plot and strip-plot complications in excipient compatibility studies.
- We explain that excipient compatibility studies are in fact mixture-process variable experiments.
- We present a flexible technology to set up highly informative experimental designs for excipient compatibility studies.
- We are the first to study I-optimal split-plot and strip-plot mixture-process variable designs.
- We illustrate our approach using a real-life excipient compatibility study.