CHARLES UNIVERSITY

FACULTY OF PHARMACY IN HRADEC KRÁLOVÉ

Department of Pharmaceutical Technology

THEORY AND APPLICATIONS OF DOE (DESIGN OF EXPERIMENTS) IN PHARMACEUTICAL TECHNOLOGY

Teorie a aplikace postupu DoE (plánovaní experimentů) ve farmaceutické technologii

Dissertation thesis

Study program: Pharmaceutical Technology

Mgr. Julie Maruška

Hradec Králové 2024

STATEMENT OF AUTHORSHIP

I hereby declare that I am the sole author of this thesis. I have written this dissertation independently under the supervision of Assoc. Prof. Erik Jurjen Duintjer Tebbens, using only the mentioned and duly cited sources and literature, and that the work has not been used in another university study programme or to obtain the same or another academic title.

In Hradec Králové Mgr. Julie Maruška

Date: _____ _____

ACKNOWLEDGMENTS

I extend my sincere gratitude to my supervisor, Prof. Duintjer Tebbens, whose invaluable guidance, endless patience, and unwavering support played a pivotal role in the completion of this dissertation. I am deeply appreciative of his exceptional support, marked by both encouragement and tolerance, especially during the challenging phases of this research work.

Special appreciation is reserved for my consultant, Prof. Šklubalova, whose hard work and dedication set an example for us doctoral students. Her expert opinion always guided us in advancing our scientific work.

Furthermore, I want to express my gratitude to two dear friends who accompanied me throughout this study, Jana Jezerska and Chiazor Ugo Ogadah. Their camaraderie, encouragement, and shared experiences were sources of strength and motivation.

On a personal note, I would like to thank my mother, who motivated me to enter postgraduate study and has been my support and assistant throughout this time. I also express my deepest thanks to my husband for his unwavering encouragement and understanding throughout this academic journey.

Last but not least, heartfelt thanks go to my son Oliver, born last year, who slept soundly while I wrote this work.

Each of these individuals has played a unique and invaluable role in shaping the outcome of this dissertation, and for that, I am profoundly grateful.

The dissertation was created with the support of the Grant Agency of Charles University (project no. 268120/2020 and 70119/2019), Specific University Research SVV 260 666 and the program for the development of scientific areas at Charles University PROGRES Q42.

ABSTRAKT

Univerzita Karlova, Farmaceutická fakulta v Hradci Králové

Katedra:	Katedra farmaceutické technologie
	Katedra biofyziky a fyzikální chemie
Kandidát:	Mgr. Julie Maruška
Školitel:	doc. Dipl.-Math. Erik Jurjen Duintjer Tebbens, Ph.D.
Konzultant:	doc. PharmDr. Zdeňka Šklubalová, Ph.D.
Název disertační práce:	Teorie a aplikace postupu DoE (plánovaní experimentů) ve farmaceutické technologii

Tradiční proces vývoje nových přípravků zahrnuje výběr kombinací různých typů faktorů ovlivňujících řadu vlastností konečné lékové formy. Tato situace je vhodná pro použití metod z oblasti statistického plánovaní experimentů (DoE). V současné době začínají nejnovější publikace v oblasti farmaceutické technologie týkající se vývoje nových lékových forem stále více začleňovat techniky plánovaní experimentů, které jsou předmětem studia v této práci.

Tato interdisciplinární disertační práce je anotovaným shrnutím publikací a výzkumných aktivit autora a klade si za cíl zkoumat postupy DoE zaměřené na jejich praktické využití v oblasti farmaceutické technologie; aplikovat vybrané techniky na reálné postupy farmaceutické technologie; a představit přehled nejúčinnějších DoE technik pro implementaci ve výzkumu v oblasti farmaceutické technologie.

V prvním publikovaném článku jsme provedli retrospektivní analýzu získaných dat ke zkoumání sypkosti čtyř frakcí farmaceutického excipientů sorbitolu, široce používaného ve farmaceutické technologii. Prozkoumali jsme vliv faktorů, jako jsou průměr otvoru násypky a velikost částic, a také jejich interakce, na hmotnostní rychlost sypání prášku. K hodnocení získaných dat jsme použili techniku přizpůsobení modelu, konstruující a hodnotící celkem patnáct modelů. Naše hlavní závěry jsou následující: (1) pro modely prokazující uspokojivou přesnost předpovědi rychlosti sypání v širokém rozsahu velikostí částic sorbitolu (0,1 až 0,346 mm) je nutné zahrnout do regresní analýzy jako faktory jak průměr otvoru, tak velikost částic; (2) plně kvadratický model je nutný pro vysokou přesnost predikce rychlosti sypání; (3) identifikovali jsme statisticky významnou interakci mezi průměrem otvoru a velikostí

částic. Tato studie, zaměřená na modelovou pomocnou látku sorbitol pro přímé lisování, ilustrovala užitečnost analýzy a modelování hmotnostní rychlosti sypání.

V druhém publikovaném článku jsme použili centrální kompozitní plán (CCD) k určení optimálních podmínek mletí excipientů za využití planetárního kulového mlýnu, jedné z nejúčinnějších technik pro zlepšení rozpustnosti omezeně rozpustných léčiv. Po výběru pěti nejvhodnějších ze dvaceti čtyř vzorků různých materiálů používaných ve farmaceutické technologii byl navržen CCD pro dva faktory (rychlost mletí a doba mletí), z nichž každý měl pět úrovní pro každou velikost mlecích koulí, se dvěma odezvami (velikost částic, distribuce velikosti částic). Pro všech deset kombinací faktorů a každou velikost koulí byl použit kvadratický model CCD na předpověď odezvy, velikosti částic. Pro tři látky z pěti byly nejlepší výsledky dosaženy při použití koulí o velikosti pět milimetrů. Přístup dosažený v této studii je považován za užitečný a očekává se, že pomůže vybrat požadované podmínky zpracování materiálu suchým mletím k dosažení výsledné velikosti částic.

Poslední publikace byla věnována vývoji liquisolid systémů, představující inovativní přístup k zlepšení rozpustnosti omezeně rozpustných léčiv. Čtyři komerčně dostupné typy koloidního oxidu křemičitého byly použity jako obalové materiály při devíti různých hodnotách R, vyjadřujících poměr hmotnosti kapaliny k hmotnosti nosiče v rozmezí od 5 do 100, k hodnocení lisovatelnosti liquisolid prášku a vlastností výlisku. Výsledky analýzy hlavních komponent (PCA) naznačují: 1) silnou pozitivní korelaci mezi výsledky úhlu skluzu a oděru tablet, které lze považovat za nejcitlivější; 2) skutečnost, že obalový materiál ovlivňuje výstupy.

ABSTRACT

Charles University, Faculty of pharmacy in Hradec Králové

Department:	Department of Pharmaceutical Technology
	Department of Biophysics and Physical Chemistry
Candidate:	Mgr. Julie Maruška
Supervisor:	Assoc. Prof. Dipl.-Math. Erik Jurjen Duintjer Tebbens, Ph.D.
Consultant:	Assoc. Prof. PharmDr. Zdeňka Šklubalová, Ph.D.
Title of dissertation:	Theory and applications of DoE (design of experiments) in pharmaceutical technology

The conventional process for developing new medicines involves selecting combinations of various types of factors that impact numerous properties of the final dosage form. This scenario is well-suited for using methods from the statistical field of design of experiments (DoE). Currently, the latest publications on pharmaceutical technology related to the development of new dosage forms are increasingly beginning to incorporate experimental design techniques, which are the subject of study in this work.

This interdisciplinary dissertation thesis is an annotated summary of the publication and research activities of the author and aims to explore the DoE approaches focusing on their practical applications within the realm of pharmaceutical technology; to apply the selected techniques in actual processes of pharmaceutical technology; and to present a review of the most useful techniques of DoE for implementation in pharmaceutical technology scientific area.

In a first published paper, we conducted a retrospective analysis of the data obtained to investigate the flow properties of four fractions of pharmaceutical sorbitol excipient, widely used in pharmaceutical technology. We explored the influence of factors such as orifice diameter and particle size, as well as their interaction, on the mass flow rate of the powder. To assess the obtained data, we utilized a model-fitting technique, constructing and evaluating a total of fifteen models. Our primary conclusions are as follows: (1) for models demonstrating satisfactory precision of flow rate prediction across a broad range of sorbitol particle sizes (0,1 to 0,346 mm), both orifice diameter and particle size need to be included as factors in the regression

analysis; (2) for highly precise prediction of mass flow rate, a fully quadratic model is necessary; (3) we identified a statistically significant interaction between orifice diameter and particle size. This study, which focused on the model excipient sorbitol for direct compression, illustrated the utility of analyzing and modeling flow rate.

In the second publication, we applied a central composite design (CCD) to determine the optimal conditions for milling excipients using a planetary ball mill, one of the most effective techniques for improving the solubility of poorly soluble drugs. After selecting the five most suitable from twenty-four samples of various materials used in pharmaceutical technology, a CCD was proposed for two factors (milling speed and milling time), each of which had five levels for each size of milling balls, with two responses (particle size, particle size distribution). For all ten factor combinations and each ball size, a quadratic response surface model was used to predict the response variable, particle size. For three substances out of five, the best results were achieved using five-mm balls. The approach achieved in this study was found to be useful and is expected to help in selecting the desired conditions of the material processing by dry milling to achieve the required particle size.

The last publication was related to the development of liquisolid delivery systems, representing an innovative approach to enhancing the dissolution of poorly soluble drugs. Four commercially available colloidal silica types were used as coating materials in nine different R values (the ratio of the mass of liquid to the mass of carrier in the range from 5 to 100), to evaluate the compressibility properties of liquisolid powder and the compact obtained. The results of PCA suggest 1) a strong positive correlation between the outcomes of the angle of slide and tablet friability, which can be considered the most sensitive outcomes; 2) the fact that the coating material does have an influence on the output.

TABLE OF CONTENTS

1 INTRODUCTION 11

2 AIMS OF THE DISSERTATION THESIS 14

3 THEORETICAL BACKGROUND 15

3.1 MODEL FITTING 20

3.1 FACTORIAL DESIGNS 27

3.1.1 FULL FACTORIAL DESIGN 29

3.1.2 FRACTIONAL FACTORIAL DESIGN 37

3.2 OTHER TYPES OF DOE USED IN PHARMACEUTICAL TECHNOLOGY 45

3.2.1 SCREENING DESIGNS 45

3.2.2 RESPONSE OPTIMIZATION 50

3.2.3 CENTRAL COMPOSITE DESIGN 52

3.2.4 MIXTURE DESIGNS 57

3.3 OTHER DESIGNS AND DATA EXPLORATION METHODS 64

4 RESULTS AND DISCUSSION 69

4.1 FLOW EQUATIONS FOR FREE-FLOWABLE PARTICLE FRACTIONS OF SORBITOL FOR DIRECT COMPRESSION: AN EXPLORATORY MULTIPLE REGRESSION ANALYSIS OF PARTICLE AND ORIFICE SIZE INFLUENCE. 69

4.2 MILLING OF PHARMACEUTICAL POWDER CARRIER EXCIPIENTS: APPLICATION OF CENTRAL COMPOSITE DESIGN 71

4.3 THE IMPORTANCE OF THE COATING MATERIAL TYPE AND AMOUNT IN THE PREPARATION OF LIQUISOLID SYSTEMS BASED ON MAGNESIUM ALUMINOMETASILICATE CARRIER. 75

5 CONCLUSION 78

6 RESEARCH OUTPUTS 80

6.1 ARTICLES RELATED TO THE TOPIC OF THE DISSERTATION 80

6.2 ARTICLES UNRELATED TO THE TOPIC OF THE DISSERTATION 81

6.3 POSTER/ORAL PRESENTATIONS 81

6.4 GRANT PROJECTS 84

6.5 SCIENTIFIC EXPERIENCE ABROAD 84

7 LIST OF REFERENCES 85

LIST OF ABBREVIATIONS

Abbreviation	Units	Name
ANOVA	-	Analysis of Variance
AP	-	Colloidal silicone dioxide (Aeroperl® 300 Pharma)
AS	-	Colloidal silicone dioxide (Aerosil® 200)
b_d	g/mL	Bulk density
bt	g/mL	Tapped density
CCD	-	Central Composite Design
CI	%	Compressibility Index
DF	-	Degree of freedom
DoE	-	Design of Experiments
EMA	-	European Medicines Agency
E_{max}	J	Total energy which is sum of the energy used for the first phase of compression, rearrangement of particles, the energy used for plastic deformation during compression and the energy used for elastic deformation after compression
FDA		U.S. Food and Drug Administration
Frac FD	-	Fractional Factorial Design
Full FD	-	Full Factorial Design
HR	-	Hausner ratio
NUS2	-	Magnesium aluminometasilicate (Neusilin® US2)
ODT	-	Orodispersible tablets
PC1	-	Principal Component 1
PC2	-	Principal Component 2
PC3	-	Principal Component 3
PCA	-	Principal Component Analysis
PEG400	-	Polyethylene glycol 400
PLS	-	Partial Least Squares
QbD	-	Quality by Design
Q_m	g/s	Mass flow rate of a powder

R^2	-	The coefficient of determination
R^2_{adj}	-	The adjusted coefficient of determination
RSM	-	Response Surface Methodology
S244	-	Colloidal silicone dioxide (Syloid® 244 FP)
S72	-	Colloidal silicone dioxide (Syloid® 72)
SD	-	Standard deviation
span	-	The width of the particle size distribution relative to x_{50}
SS_E	-	The error sum of squares
SSLF	-	Lack of fit sum of squares
SSPE	-	Pure error sum of squares
SS_R	-	The regression sum of squares
SS_T	-	The total sum of squares
x_{10}	µm	Particle size corresponding to 10% of the cumulative frequency
x_{50}	µm	Particle size corresponding to 50% of the cumulative frequency
x_{90}	µm	Particle size corresponding to 90% of the cumulative frequency

1 INTRODUCTION

Pharmaceutical production demands considerable efforts to optimize the critical attributes of materials and process parameters. The pharmaceutical industry has been operating for many decades utilizing a one factor at a time and one variable at a time approach or changing one separate variable at one time point. These methods, however, are not renowned for consistently yielding satisfactory outcomes in terms of product robustness and performance. Pharmaceutical drug products are complex entities that include many factors that contribute to the quality, safety, and efficacy.

While traditional optimization tools can provide workable solutions, they require a significant amount of time, effort, and resources. Moreover, due to the complexity, it becomes extremely challenging to consider and optimize all product and process-related factors simultaneously. Traditional approaches would necessitate a very high number of experiments to do so. Additionally, recognizing the presence of interactions among factors is crucial, as these interactions significantly influence product quality.

Design of Experiments (DoE) is one of the systematic tools that is useful for improving of product quality, robustness, and process performance. It specifically focuses on product and process development, guided by set objectives. DoE employs robust statistical principles to carry out a minimal number of experiments while generating maximum information. It contributes to significant savings by reducing the time, effort, and resources required for conducting experiments. Additionally, it enhances understanding of potential interactions between input factors. Using DoE, one is also capable to predict formulation performance and identify errors through understanding interaction effects among factors. These benefits are not usually observed with traditional approaches (Beg et al., 2019; Beg, 2021).

The origin of DoE can be laid at the mid-20th century when the concept of experimental design was introduced in the field of statistics. At the end of the 50s of the 20th century, statisticians such as Ronald Fisher and Frank Yates developed the theory of experimental design (Box, 1922; Sprent, 1973), which provided a systematic approach for planning, conducting, analysing, and interpreting experiments. This approach was quickly adopted by scientists in various field, including pharmaceutical sciences, for experiments optimization and gaining more insights into the processes (for example to study the effects of different formulation parameters on drug release

rates, to optimize drug stability and shelf life or to develop more efficient production processes) (Politis et al., 2017).

Over time, regulatory agencies like the EMA and FDA started acknowledging the significance of DoE in the pharmaceutical industry. They began requiring the incorporation of DoE in drug development and production processes[1].

Today, DoE is a part of many processes in pharmaceutical technology, including:

1. Formulation development: DoE can be used to study the effects of different formulation parameters, such as the type and quantity of excipients, on drug release rates, bioavailability, and stability (Shariare et al., 2019).
2. Process optimization: DoE can be used to optimize manufacturing processes such as blending, granulation, and drying, to improve product quality and consistency, reduce costs, and minimize waste (Jaydip et al., 2020; Liu et al., 2020; Mamidi et al., 2021; Zidan et al., 2022).
3. Analytical method development: DoE can be used to optimize analytical methods such as chromatography and spectroscopy, to improve sensitivity, accuracy, and precision (El-Sayed et al., 2023).
4. Quality control: DoE can be used to develop robust quality control procedures, to ensure that drug products meet regulatory requirements and are safe and effective for patients (Huang et al., 2020).

Some of the benefits of using DoE in pharmaceutical technology include:

i. Improved process understanding compared to conventional methods: DoE can help researchers and technologists to gain a deeper understanding of the factors that affect drug development and manufacturing processes, which can lead to more efficient and effective processes.

ii. Faster development times relative to traditional methods: DoE can help to identify optimal process conditions requiring less experiments, which can reduce development times.

iii. Reduced costs in comparison to traditional approaches.: DoE can help to optimize processes and reduce waste, which can lead to significant cost savings for pharmaceutical companies.

[1] https://www.ema.europa.eu/en/human-regulatory-overview/research-and-development/quality-design

iv. Improved product quality: DoE can help to identify critical process parameters and control limits, which can improve product quality and consistency and reduce the risk of product failures or recalls.

As for the structure of the thesis, the theoretical part will explore the primary directions of experimental design, specifically:

In the third chapter, we will focus on the most classic factorial designs (full and fractional FD). Additionally, we will consider other popular types of designs used in pharmaceutical technology based on the goals of scientific work, such as screening designs, optimization approaches, and mixture designs. Finally, we will explore methods for handling large amounts of acquired data.

The fourth chapter will examine practical examples of the application of certain DoE techniques in pharmaceutical technology.

As the conclusion of this work, the author's scientific results will be summarized.

All the calculations were carried out, unless stated otherwise, using Microsoft Excel, provided by Microsoft Corporation in 2018. The software is available at: https://office.microsoft.com/excel.

2 AIMS OF THE DISSERTATION THESIS

Pharmaceutical technology has several characteristics that make the development process highly challenging, requiring a tailored design of experiments (DoE). Among these challenges are the complexity of the formulation since dosage forms often involve intricate formulations, such as drug delivery systems, multiple active ingredients, or controlled release mechanisms. Additionally, there is the necessity to regulate pharmaceutical products, with each medicine required to demonstrate its safety, efficacy, and quality. Robustness requirements of the manufacturing process and quality control further add to the complexity. Moreover, there are limitations regarding the availability and cost of resources and materials used in pharmaceutical production.

This interdisciplinary work aims to review modern approaches to the design of experiments, specifically focusing on their practical applications within the realm of pharmaceutical technology while considering its specific characteristics. The main goal is to create a manual for practitioners on how to choose the right experimental design to solve problems in the technological development of various dosage forms.

The primary objectives of this thesis can be outlined as follows:

i. A review of the modern design of experiments techniques

ii. Experimental application of selected techniques in the pharmaceutical development of solid dosage forms (resulting in publications)

iii. Evaluation of the obtained results, making the relevant conclusions based on statistical significances and indicators of model quality and recommendations for the implementation of the selected design of experiments in pharmaceutical technology (included in publications).

3 THEORETICAL BACKGROUND

For a proper understanding of a designed experiment, it is essential to have a good understanding of the process. A system or process can be conceptualized as a combination of parameters or factors that convert an input into an output consisting of one or more measurements or having one or more observable property y, referred to as response variables (see Figure 1). Some of the process variables or factors denoted x_1 to x_k, can be controllable, while others, z_1 to z_q, are uncontrollable and known as nuisance factors (Montgomery, 2017).

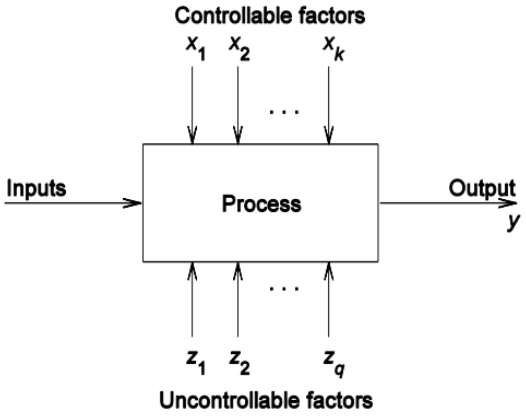


Figure 1. General model schematic of a process or system (Montgomery, 2017), (Antony, 2014).

In DoE the following types of variables are distinguished:

- Input variables (factors): the independent variables that are intentionally manipulated by the experimenter to study their effects on the response variable(s). Factors can be further categorized into:
  a. *Quantitative variables*: continuous variables that can take any value within a specific range (e.g., speed of rotation or concentration).
  b. *Qualitative variables*: categorical variables that do not take numerical values (e.g., equipment type, excipient brand or operator).
- Response variables (output): these are the dependent variables that are measured or observed because of the changes in the input variables. The objective of DoE is to understand the relationship between the factors and the response variables.
- Control variables: these are the variables that may influence the response but are not the focus of the experiment. Control variables are held constant throughout

the experiment to prevent their influence on the response variable(s) (for example: temperature in a laboratory, pressure, relative humidity during storage, equipment settings or a certain period during which the experiment takes place).

- Nuisance variables: these are the variables that can introduce variability or noise into the experiment but are not of primary interest. These variables may be uncontrollable, and their effects are minimized through randomization, replication, and blocking techniques (Beg et al., 2019). For example, raw material batch, season, time of the day, operator etc.
- Confounding variables: these are variables that may influence both the input variables (factors) and the response variables, making it difficult to determine the true relationship between the factors and the response. Confounding variables need to be controlled, if possible (then they are in fact, controlled variables) or accounted for in the experimental design and analysis to draw valid conclusions. For example, the accuracy of tablet press calibration, which affects the compression of tablets (input) to the correct hardness, can have an impact on the dissolution rate of the tablets (output). Additionally, the purity of pharmaceutical ingredients can significantly influence the overall quality of the final product.

By deliberately altering input process variables (or factors), the corresponding changes in the process output can be observed. In the case of new product development, design parameters are modified to ensure the design performance is less susceptible to all sources of variation (Montgomery, 2017). The insights gained from well-planned, executed, and analysed experiments can contribute to enhanced product functionality, reduced scrap or rework rates, shorter product development cycles, decreased variability in production processes, increased process throughput yields, and improved process capabilities.

Besides process understanding, four core principles of experimental design should be also mentioned, namely: randomization, replication, blocking, and control (Antony et al., 2011). These principles aim to minimize or eliminate experimental bias. It is crucial to recognize that substantial experimental bias may lead to incorrect optimal settings or, in some instances, obscure the influence of truly significant factors. Below is a more detailed description of each of these principles.

- Randomization: Randomization involves the random assignment of experimental units to different treatment groups. The aim is to eliminate or minimize the influence of confounding variables, nuisance variables, or systematic errors. The random process ensures that every possible allocation of treatments has an equal probability.
- Replication: Replication involves conducting multiple runs of the experiment under identical conditions. The purpose is to estimate experimental error and enhance the reliability of the results. Replication helps accommodate the inherent variability in the process, providing a more accurate estimate of the treatment effects.
    - It allows the experimenter to obtain an estimate of the experimental error.
    - More replications would provide an increased precision by reducing the standard error (SE) of the sample mean which is defined as $s_{\bar{y}} = \frac{s}{\sqrt{r}}$, where s is the sample standard deviation and r is the number of replications (Casler, 2015).
    - In linear regression, replicates allow for a lack-of-fit test (described later).
- Blocking: Blocking involves grouping experimental units with similar characteristics together and applying treatments within these groups (blocks). Blocking must be appropriately implemented in the statistical analysis. Blocking helps isolate the effect of the treatment from other sources of variation and improve the precision of the experiment.
- Control: Including a control group in the experiment is known as control. This is done to provide a baseline for comparison with the treatment groups. The inclusion of a control group helps determine whether the treatment has a significant effect on the response variable. Any differences observed between the control and treatment groups can be attributed to the treatment itself.

For better understanding all these principles consider, as an example, the optimization of a tablet manufacturing process. One of the main parameters of the tableted mixture is its flow properties. To measure mass flow there are various

methods but let us focus here on the most classic, the gravimetric method. In the case of the gravimetric method, the flow rate through an orifice is measured as the mass per time flowing from a hopper. The main two independent factors that impact the mass flow rate are bulk density of the powder and orifice diameter of the hopper. To minimize bias and the impact of unknown variables, one should conduct the experimental runs (each combination of powder with different bulk density and orifice diameter of the hopper) in a random order. For instance, if there were four levels of bulk density of the tablet blend and four orifice diameters (a total of 16 experimental runs), these could be conducted in an order determined by a random number generator. To estimate the inherent variability in the process and to improve the precision and reliability of the results, the technologist can repeat each experimental run multiple times. For instance, each combination of bulk density of the mixture and orifice diameter might be replicated ten times. If there are known nuisance variables that could affect the response but are not of interest, these can be controlled through blocking. For example, if the measurements must be carried out on two different equipment that might not operate identically, the technologist can use blocking to control for this. And finally, to check the stability and consistency of the process, control samples (measurements could be made using a standard orifice diameter with each bulk density) could be included in the experiment. If the control samples produce results within the expected range, this provides confidence that the process is functioning as expected. If the control samples produce unexpected results, this might indicate an issue with the process that needs to be addressed.

A general DoE procedure consists of the following steps (Gujral et al., 2018; Politis et al., 2017):

i. Define the objective: the first step in DoE is to clearly define the problem and identify the objectives of the experiment. This involves identifying the process or system to be studied, all input factors that are involved in the process, and the output responses that are of interest.
ii. Define the variables (factors/responses): the next step is to identify and select the specific factors that are most likely to affect the output responses, while disregarding those that are not important.

iii. Select the design of experiment: the design should consider the number of factors, the type of factors (quantitative or qualitative) and the available number of levels of each factor.

iv. Conduct the experiment (data collection): the experiment should be conducted according to the designed plan. This involves collecting data on the input factors and output responses, controlling for any extraneous variables that could affect the results, and randomizing the order of the experimental runs.

v. Analyse the data: the data collected from the experiment is analysed using statistical methods to determine the effects of each factor on the response variable. This could involve techniques such as analysis of variance (ANOVA), regression analysis, or optimization techniques.

vi. Interpret the results: the results of the analysis are interpreted to identify the critical factors that have a significant effect on the response variable, and the optimal settings or levels of these factors to achieve the desired response.

vii. Validate the results: finally, the results of the experiment are verified and validated by conducting further experiments, if necessary, to ensure that the results are robust and repeatable.

There is a number of software packages available for implementation of DoE: Design-Expert® (tailored for DoE), JMP®, Statistica®, Minitab®, OPTIMA, CAMO, R, SigmaXL, Centurion, and ReliaSoft which usually provide interface guide at every step during the entire product development cycle. Software is also available for chemometric analysis through multivariate techniques like Multi-Normal Linear Regression Analysis (MNLRA), Principal Component Analysis (PCA), PLS (Partial Least Squares), etc. which include MODDE®, Unscrambler®, SIMCA®, CODDESA®. For performing quality risk management, software like Minitab®, Risk®, Statgraphics, FMEA-Pro, iGrafx, etc., can be used (Dhoot et al., 2019).

3.1 MODEL FITTING

In most scientific experiments, there is a connection between two or more variables, and understanding and mathematically modelling this relationship is beneficial for the study. For the moment, we assume we have a single response or dependent variable Y that relies on k independent variables, denoted x_1, x_2, ..., x_k. The relationship can be characterized by a mathematical technique known as regression, which represents a simplification of the real-world behaviour to a mathematical model. The main principle of regression techniques is to estimate and model the relationship between variables by minimizing the distance between observations and model predicted expected value. As the exact functional relationship between variables is typically unknown, low-order polynomial regression models are frequently employed as approximating functions. Regression techniques are also often employed in the analysis of data from unstructured experiments, like those arising from observations of uncontrolled phenomena or historical records. Moreover, these techniques prove to be highly valuable in scenarios where a structured experiment has experienced unexpected issues or deviations and in post experimental data assessment (Hinkelmann, 2007).

In this section, we will illustrate fitting a multiple linear regression model to available historical data. The multiple linear regression model is a regression model with more than one independent variable and could be represented as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ \ldots \ + \beta_k x_k \ + \varepsilon \qquad (1)$$

where Y represents the dependent variable or response, x_1, x_2 ... x_k represent independent variables or predictor variables or regressors, β_0 is the intercept and β_j, $j = 1, 2, \ldots, k$ is called regression coefficient or slope for the given variable, ε represents the error term or residual.

To estimate the coefficients $\beta_0, \beta_1, \ldots \beta_k$ in a multiple linear regression model, the most used method is the method of least squares. Consider a study with n observations (where $n > k + 1$) with response values Y_i ($i = 1, 2 \ldots n$), where the corresponding value of the jth regressor variable at Y_i will be denoted as x_{ij}. The input data for multiple linear regression can be organized as in Table 1.

Table 1. Organization of input data for multiple linear regression for the factors $x_1 \ldots x_k$.

***Y* (response)**	**x_1 (factor 1)**	**x_2 (factor 2)**	**…**	**x_k (factor *k*)**
Y_1	x_{11}	x_{12}	…	x_{1k}
Y_2	x_{21}	x_{22}	…	x_{2k}
…	…	…	…	….
Y_n	x_{n1}	x_{n2}	…	x_{nk}

We consider the example discussed before with the following logarithmized values (see Table 2). The values were logarithmized to be able to apply multilinear regression (for more details, see (Marushka et al., 2022)).

Table 2. Example of logarithmized values for factors x_1 and x_2 and outcome Y.

Mass flow rate (g/s)	***Y***	**Bulk density (g/mL)**	**x_1**	**Orifice diameter (mm)**	**x_2**
1,96	0,67	0,59	-0,53	6,00	1,79
5,37	1,68	0,61	-0,49	8,00	2,08
10,33	2,34	0,62	-0,48	10,00	2,30
27,52	3,31	0,64	-0,45	15,00	2,71

The regression model in matrix notation can be expressed as:

$$Y = X\beta + \varepsilon \qquad (2)$$

where

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, X = \begin{bmatrix} 1 x_{11} x_{12} \ldots x_{1k} \\ 1 x_{21} x_{22} \ldots x_{2k} \\ \vdots \; \vdots \; \vdots \; \ldots \; \vdots \\ 1 x_{n1} x_{n2} \cdots x_{nk} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \text{ and } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Y is the outcome vector (n x 1), X is a matrix representing the corresponding values of the independent variables with dimensions (n x p), where $p = k + 1$ is the number of coefficients and β is a column vector representing the estimated regression coefficients with dimensions (p x 1).

In the context of multiple linear regression, there are specific requirements, namely: the errors ε in the model are normally distributed with mean zero and with the same variance σ^2 for every observation, and the observations Y_i are normally and independently distributed with mean $\beta_0 + \sum_{j=1}^{k} \beta_j x_{ij}$ and variance σ^2.

Using the n observations Y_i we can compute estimates $\hat{\beta}_0$, $\hat{\beta}_1$, … $\hat{\beta}_k$ of the coefficients β_0, β_1, ... β_k for the model (1) supposed to hold for the population of all observations. Applying the least-squares method, we obtain the estimates as:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = (X^T X)^{-1} X^T Y,$$

where X^T is the transpose of the matrix X, and $(X^T X)^{-1}$ is the inverse of the $X^T X$. We assume the inverse to exist, i.e. X has full column rank, and there is no collinearity.

In our example we have:

$$Y = \begin{bmatrix} 0{,}67 \\ 1{,}68 \\ 2{,}33 \\ 3{,}31 \end{bmatrix}, X = \begin{bmatrix} 1 & -0{,}53 & 1{,}79 \\ 1 & -0{,}49 & 2{,}08 \\ 1 & -0{,}48 & 2{,}30 \\ 1 & -0{,}45 & 2{,}71 \end{bmatrix}, \text{hence}$$

$$\hat{\beta} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 \\ -0{,}53 & -0{,}49 & -0{,}48 & -0{,}45 \\ 1{,}79 & 2{,}08 & 2{,}30 & 2{,}71 \end{bmatrix} \begin{bmatrix} 1 & -0{,}53 & 1{,}79 \\ 1 & -0{,}49 & 2{,}08 \\ 1 & -0{,}48 & 2{,}30 \\ 1 & -0{,}45 & 2{,}71 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 \\ -0{,}53 & -0{,}49 & -0{,}48 & -0{,}45 \\ 1{,}79 & 2{,}08 & 2{,}30 & 2{,}71 \end{bmatrix} \begin{bmatrix} 0{,}67 \\ 1{,}68 \\ 2{,}33 \\ 3{,}31 \end{bmatrix}$$

$$= \begin{bmatrix} 6{,}35 \\ 15{,}68 \\ 1{,}48 \end{bmatrix}.$$

The final mathematical model will look as follows:

$$\hat{Y} = 6{,}35 + 15{,}68 \cdot x_1 + 1{,}48 \cdot x_2$$

The difference between the actual observed value Y_i and the corresponding model predicted value $\hat{Y}_i$ is the residual: $e_i = Y_i - \hat{Y}_i$.

Table 3. The actual observations Y_i, the predicted or fitted values $\hat{Y}_i$ and the residuals.

Y_i	x_1	x_2	$\hat{Y}_i$	Residual (e_i)
0,67	-0,53	1,79	0,68	0,01
1,68	-0,49	2,08	1,70	0,02
2,34	-0,48	2,30	2,24	0,10
3,31	-0,45	2,71	3,34	0,03

For estimation of σ^2 (the population variance of the data around the predicted expected value), consider

$$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = Y'Y - \hat{\beta}'X' \qquad (3)$$

SS_E is called error or residual sum of squares. An unbiased estimate of σ^2 with $n - p$ degrees of freedom (number of independent values which can vary without violating any given constraints) can be calculated by:

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} \qquad (4)$$

To test whether a linear relationship exists between the dependent variable *Y* and a subset of the independent or regressor variables $x_1, x_2, \ldots, x_k$ several approaches can be employed. The most popular is a hypothesis testing approach for which the main idea is to assess the *p*-value and, reject the null hypothesis if the *p*-value is less than α (for more details, see (Myers et al., 2016)). Conducting *t*-tests for each coefficient β_j to estimate the effect of a particular predictor variable on the outcome variable may be useful. If the *p*-value for a coefficient is above the chosen significance level (commonly $\alpha = 0{,}05$), then one cannot reject the null hypothesis, suggesting that the true coefficient for the population of all observations is not statistically significantly different from zero, and the predictor variable does not have a significant effect on the outcome. But if the *p*-value is below 5% that means that the coefficient is statistically significant, and the predictor variable has a significant effect on the outcome. By conducting *t*-tests for each coefficient, one can assess which predictor variables are contributing significantly to the model and which ones may not be useful for explaining the outcome variable. This helps in model selection and interpretation.

The examples with calculations (see Table 9) also include the coefficient of determination R^2:

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_E}{SS_T} \qquad (5)$$

where SS_E - the error sum of squares;

SS_R – the regression sum of squares, $SS_R = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2$, where $\bar{Y}$ represents the mean of the observed values of *Y*.

SS_T – the total sum of squares, which represents the total variation of the observed values of the dependent variable (*Y*) from their mean and is calculated as:

$$SS_T = \sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

where $\bar{Y}$ represents the mean of the observed values of *Y* (see Figure 2).

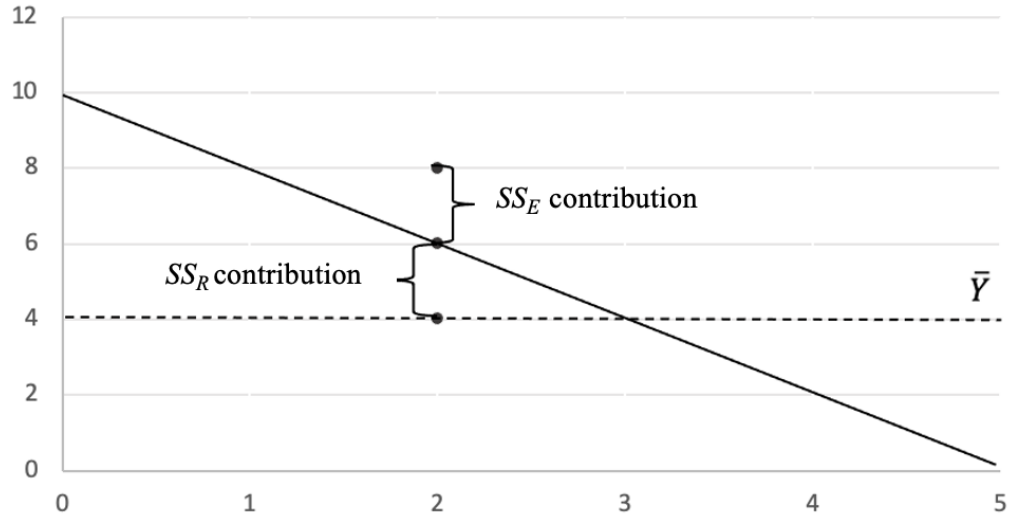


Figure 2. Visual representation of the contribution to the total error of the SS_E and SS_R for the particular sample.

R^2 measures the extent to which the variability of Y is explained when the regressor variables x_1, x_2, ..., x_k are incorporated into the model. However, a high R^2 value does not automatically indicate high quality of the regression model. The inclusion of an additional variable in the model will always increase R^2, regardless of whether that additional variable is statistically significant or not. Consequently, models with R^2 values close to one can still yield inaccurate predictions for new observations or provide imprecise estimates of the mean response.

Because of the property of R^2 to increase with the addition of terms to the model, it is sometimes reasonable to use an adjusted R^2 statistic, which is defined as follows:

$$R^2_{adj} = 1 - \frac{SS_E/(n-p)}{SS_T/(n-1)} = 1 - \left(\frac{n-1}{n-p}\right)(1 - R^2)$$

Typically, the adjusted R^2 value does not exhibit a consistent increase as additional variables are included in the model. When unnecessary terms are added, the value of the adjusted R^2 statistic often decreases (Montgomery, 2017; Tošenovský, 2010).

Besides R^2 and R^2_{adj}, the quality of the model can be measured, among others, by the precision of prediction, the variability of coefficient estimates, and the prediction variances (Zatloukal et al., 2012). Prediction precision refers to the accuracy

and reliability of the predictions made based on the experimental data and statistical models. It is a measure of how closely the model's predictions match the actual outcomes or responses in each experiment.

The precision of the prediction (Δ) is the size of the relative difference between the value measured and that predicted:

$$\Delta\,(\%) = ((Y_i - \hat{Y}_i)/Y_i) \cdot 100\%.$$

The variability of coefficient estimates in statistical models refers to the extent to which the values of the coefficients ($\hat{\beta}$) can vary across different samples or datasets. This variability reflects the uncertainty inherent in estimating population parameters based on a finite set of observations. An important property of the least squares estimator is that it produces an unbiased estimator of β, meaning that the expected value for the estimator is the true value of the parameter β. An unbiased estimator is desirable because it, on average, does not systematically overestimate or underestimate the true parameter value (Myers et al., 2016):

As for the variance of $\widehat{\beta}$, the covariance matrix of $\hat{\beta}$ is a $p \times p$ symmetric matrix whose (j, j)th element is the variance of $\hat{\beta}_j$ and whose (i,j)th element is the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$. The covariance matrix of $\hat{\beta}$ is:

$$\text{Cov}\,(\hat{\beta}) = \sigma^2(X^TX)^{-1}$$

This holds for any model matrix in a regression model of the form (2), including models with nonlinear terms. σ^2 is often estimated using (4).

A third characteristic of the quality of the design, is the prediction variance. In any process of constructing a model, the primary sampling characteristic is the variance of the predicted response at a specific point of interest x, typically represented as (Myers et al., 2016):

$$\hat{y}(x) = x^{(e)T}\widehat{\beta}$$

where $x^{(e)}$ is a function of the location at which one is predicting the response; the e indicates that $x^{(e)}$ is just x expanded to model space. For example, for 2 factors and a second-order model without interaction we have $x^{(e)T} = [1, x_1, x_2, x_1^2, x_2^2]$ and $\hat{\beta}^T = [b_0, b_1, b_2, b_{11}, b_{22}]$. Under the independent and identically distributed error assumptions and with the assumption of homoscedasticity (the constant σ^2) we have for the prediction variance:

$$Var[\hat{y}(x)] = \sigma^2 x^{(e)T}(X^TX)^{-1}x^{(e)}$$

where σ can be estimated through (4).

The predicted variance of predicted value can provide an idea about the relative quality of the predicted response values in various locations in the design region.

Linear regression was the main tool used to study the influence of the orifice diameter size and particle size on the flow properties of pharmaceutical sorbitol. The results were published as a first-author publication (Marushka et al., 2022):

MARUSHKA J., HURYCHOVÁ H., ŠKLUBALOVÁ Z., DUINTJER TEBBENS J.: Flow Equations for Free-Flowable Particle Fractions of Sorbitol for Direct Compression: An Exploratory Multiple Regression Analysis of Particle and Orifice Size Influence. *Pharmaceutics*, 2022, 14(8):1653.
DOI: 10.3390/pharmaceutics14081653. ISSN: 1999-4923, IF_{2022} 5.4, Q_{AIS} 2

These results were also presented at the following international conferences:

MARUSHKA J., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: On linearized flow equations to assess the flow rate of sorbitol for direct compression. 3rd IMA Conference on Dense Granular Flows - IMA, Cambridge, UK 01 - 04. 07. 2019

MARUSHKA J., DUINTJER TEBBENS J., ŠKLUBALOVA Z.: Determination of the significance of the influence of interaction between particle size and office diameter for the prediction of a flow rate of pharmaceutical filler for direct compression. X Congress of the Slovak Pharmaceutical Society, Bratislava, Slovakia, 05 – 07.09.2019.

3.1 FACTORIAL DESIGNS

To systematically study the effect of numerous factors and their interactions on a particular outcome simultaneously, it is beneficial to employ so called factorial experiments, also known as factorial experimental designs or arrays. A factorial design includes two or more input variables, each with distinct values or "levels". If all factors have the same number of levels and all possible combinations of levels are evaluated experimentally, these designs are denoted as a X^k design, according to the total number of experimental runs (without replicates), with X representing the number of levels and k indicating the number of factors (Beg et al., 2019).

A general factorial design allows the experimenter to roughly understand the effect of factors and often as well interactions between them. The influence of a factor on the response is characterized by the variation in the outcome prompted by a modification in the factor's level. This is often referred to as the main effect, given that it pertains to the principal factors under examination in the experiment. In certain experiments, it might occur that the change in output, resulting from the differing levels of one factor, does not remain consistent across all levels of other factors. When such a scenario arises, it indicates that there is an interaction between these factors (Figure 3).

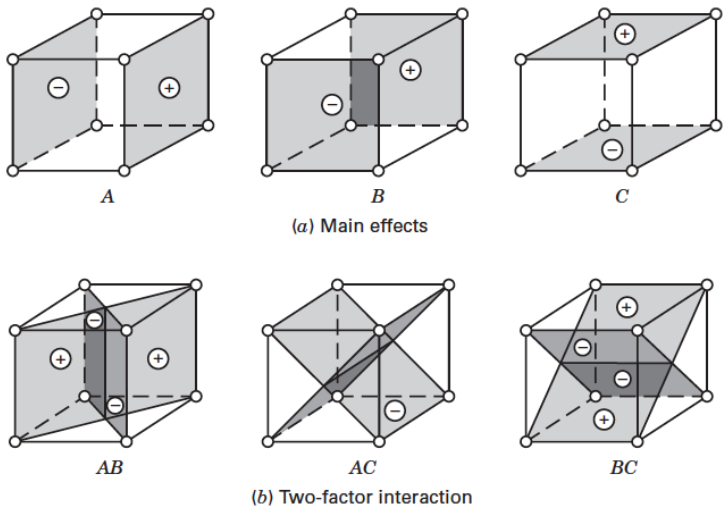


Figure 3. Visual representation of the main effect and two-factor interaction between three factors (A, B and C) (Montgomery, 2017).

To estimate the main effect of the factor *A* (or *B, C* see Figure 3, (a)) one can use the average difference method, in which the average at the low level ($\bar{A}_-$) is subtracted from the average at the high level ($\bar{A}_+$)

$$ef(A) = (\bar{A}_+ - \bar{A}_-)/2.$$

If the changes in the level of factor *A* result in different changes in the value of the response variable for different levels of factor *B* (see Figure 3, (b)), one can say that there is an interaction effect between the factors.

The interaction effect between two factors *A* and *B* is the average difference between effect of *A* at the low level of factor *B* and in effect of *A* at the high level of factor *B:*

$$ef(AB) = 1/2(ef(A)/\ B_+ - ef(A)/\ B_-) = 1/2((\overline{A_+B_+} - \overline{A_-B_+} - (\overline{A_+B_-} - \overline{A_-B_-}))) =$$

$$1/2(\overline{A_+B_+} + \overline{A_-B_-} - \overline{A_-B_+} - \overline{A_+B_-}).$$

Unlike in a simple regression model

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

the regression model with interaction is expressed through adding a multiplicative term:

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

3.1.1 FULL FACTORIAL DESIGN

The most used factorial experimental designs in pharmaceutical industry are full factorial (full FD) and fractional factorial designs (frac FD).

A full factorial design is a specific type of factorial experiment that tests all possible combinations of the considered levels of the given factors. This method, which considers the main effects of all factors and their interactions (including interaction between more than two factors), provides a deeper understanding of the system's behaviour compared to other experimental designs. Full FD models the response with an intercept, a term linear for every factor, and possibly interaction terms, but without quadratic terms for the individual factors.

Two-level full factorial designs are the most basic and used form of factorial designs. They possess advantageous orthogonality (which will be explained later in this section) and yield a manageable number of experiments, making them suitable for both factor screening and factor optimization studies. When investigating *k* factors at just two levels, the number of experimental runs is calculated as two to the power of the number of factors. Conversion of original variables to coded variables at the levels high (1) and low (-1) can be calculated using the following formula:

$$x^c = \frac{x_o - \frac{x_{max} + x_{min}}{2}}{\frac{x_{max} - x_{min}}{2}}$$

where x_o refers to a variable x in its natural units, x^c represents the coded variable, x_{max} and x_{min} represent the higher and the lower level of x respectively (Tošenovský, 2010). The result of this formula will be a coded value of either 1 or -1 for x_{max} or x_{min}, respectively. Coded variables simplify experimental design and analysis in DoE by standardizing and centering the variables, making it easier to compute, understand and interpret the results, standardizing computations, with the involved matrices facilitating efficient and orthogonal designs, and enhancing the robustness and precision of the analysis.

Two-level factorial designs are of special importance for several reasons:

- they require relatively few runs per factor studied;
- the analysis of the results derived from these designs can largely be conducted through basic calculations, and computer-generated visualizations;

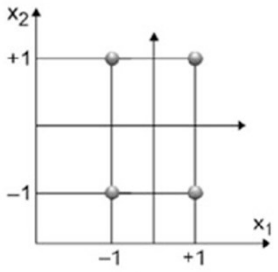
- these designs can be appropriately expanded when a more detailed local investigation is necessary;
- they also lay the foundation for two-level fractional factorial designs, that are discussed later, where only selected runs of the full factorial design are carried out. These fractional designs are especially useful for the purpose of screening factors that will be discussed in subsection 3.3.1. (Box et al., 2005).

Typical design matrices of 2^k designs with visual illustration are depicted in Table 4 (Das et al., 2023).

Table 4. 2^2 factorial design with two factors (x_1 and x_2) at two levels (1 and -1) and 2^3 factorial design with three factors (x_1, x_2 and x_3) at two levels (1 and -1).

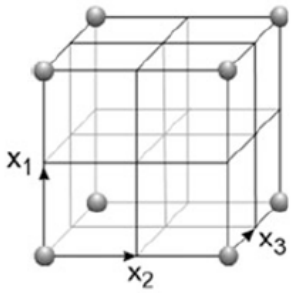
2^2 factorial design

Experimental run	x^c_1	x^c_2
1	-1	-1
2	1	-1
3	-1	1
4	1	1



2^3 factorial design

Experimental run	x^c_1	x^c_2	x^c_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



Some properties of the 2^k factorial design matrix are:

- every column displays an equal number of -1 and 1, representing low and high settings, respectively.
- orthogonality refers to the property where any two columns of the design matrix (representing two different factors or interactions) are uncorrelated, meaning that the sum of the products of their coded levels is zero: $\sum_{i=1}^{2^k} x_{ij} \cdot x_{il} = 0$ for the jth and lth columns. Orthogonality allows a clear

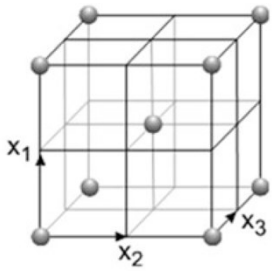
assessment of the individual impact of each factor on the response variable without interference from other factors (Oimoen, 2019).

Here one should also dwell on one of the shortcomings of these designs and its solution. To test the assumption whether the factors have linear effects on the response, one can add to the original design central points (Table 5). Adding centre points to the initial design or as an enhancement post the initial testing can be beneficial, especially when a linear model fails to provide a good fit.

Table 5. 2^3 factorial design with three factors (x_1, x_2 and x_3) at two levels (1 and -1) and added center points (0).

2^3 factorial design and added center points

Experimental run	x^c_1	x^c_2	x^c_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	0	0	0
10	0	0	0
11	0	0	0



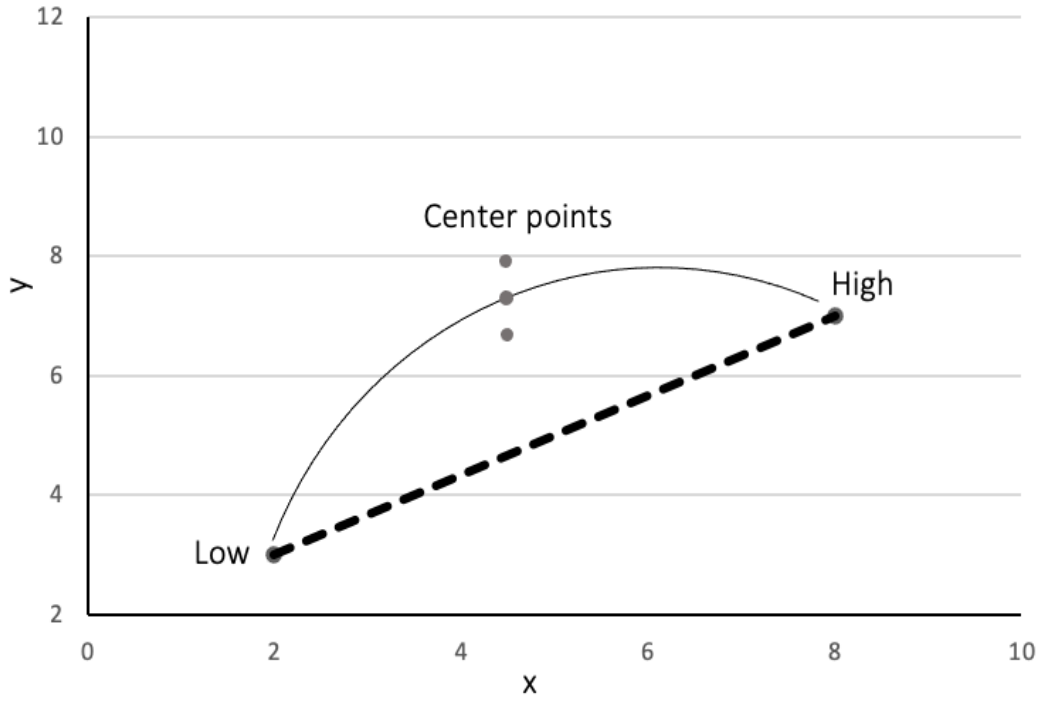
Incorporating centre points into a two-level design aids in estimating the pure error, as well as testing a lack-of-fit. Pure error is the variability in measuring the dependent variable, forming one part of the residual error. By subtracting the pure error from the residual error, one can estimate how well or poorly the model fits the data[2]:

$$\underbrace{\sum_i \sum_j (Y_{ij} - \hat{Y}_i)^2}_{SS_E} = \underbrace{\sum_i \sum_j (\bar{Y}_i - \hat{Y}_i)^2}_{SSLF} + \underbrace{\sum_i \sum_j (Y_{ij} - \bar{Y}_i)^2}_{SSPE}$$

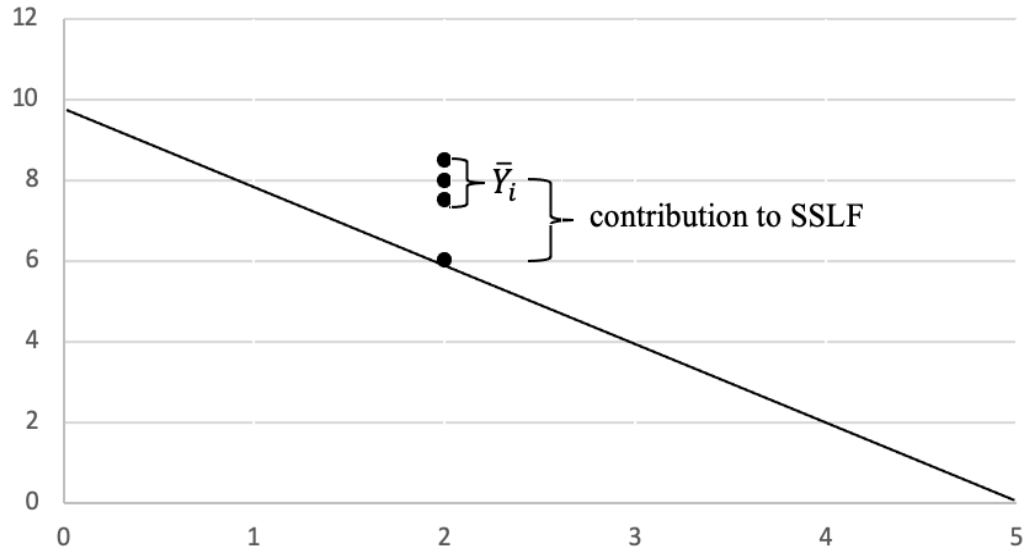
SS_E (error sum of squares) $SSLF$ (Lack of fit sum of squares) $SSPE$ (Pure error sum of squares)

[2] https://online.stat.psu.edu/stat462/node/111/

where $\bar{Y}_i$ represents the average of all of the observed *Y* values at the *i*th *x*-value (see Figure 4B, where the number of observations is three), Y_{ij} denotes the *j*th measurement made at the *i*th *x*-value in the data set and $\hat{Y}_i$ is the predicted response for the *i*th *x*-value.



A.



B.

Figure 4. Visual representation of the lack of fit for the particular sample (A); and the contribution to *SSLF* of the average of all the observed values for a particular sample (B).

The lack-of-fit test compares the average response of the centre points with the predicted average response of the factorial points for some factor (see Figure 4). A considerable discrepancy between those values signified that the linear model is inadequate, hence showing a substantial lack-of-fit. When this occurs, it suggests that the model is failing to adequately capture the variations in the response. This indicates that there are aspects of the response that are not being accounted for by the current model (Beg et al., 2019; Oimoen, 2019).

Typically, the number of centre points is chosen to be between 3 and 5. However, the decision to use centre points depends on the specific factorial design being employed and the objectives of the experiment. Centre points are used in full factorial design when the goal is to optimize factor values. On the other hand, centre points are usually omitted in fractional factorial designs (explained in section 2.2.2) due to their low-resolution structure (resolution will be explained later in the next subsection) and the overall aim to minimize the total number of experimental runs (Beg, 2021).

The application of a full factorial design in pharmaceutical technology could be demonstrated using the data from the previous chapters. Let us consider an experiment which involves three independent factors each with two levels (2^3), namely orifice diameter, bulk density, and relative humidity (Table 7). Each of these factors was examined at two levels coded as a low level (-1), and a high level (1), which gives a 2^3 full factorial design, with eight experimental runs (see Table 6).

Table 6. The natural and coded values for the factors bulk density (x_1), orifice diameter (x_2) and relative humidity (x_3), each with two levels.

Bulk density (g/mL) x_1	x^c_1 **(coded)**	**Orifice diameter (mm)** x_2	x^c_2 **(coded)**	**Relative humidity (%)** x_3	x^c_3 **(coded)**
0,588	-1	6	-1	30	-1
0,639	1	15	1	60	1

Table 7. 2^3 The full factorial design matrix.

Run	x^c_1	x^c_2	x^c_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

Table 8. The natural values for the factors x_1 (bulk density), x_2 (orifice diameter), x_3 (relative humidity) without interaction with the outcome mass flow rate (Y).

Run	x_1	x_2	x_3	Y	***SD****
1	0,588	6	30	1,96	0,01
2	0,639	6	30	2,41	0,17
3	0,588	15	30	2,37	0,01
4	0,639	15	30	16,65	0,32
5	0,588	6	60	2,12	0,05
6	0,639	6	60	2,51	0,20
7	0,588	15	60	2,70	0,13
8	0,639	15	60	16,87	0,37

**replicated 10 times*

**SD (standard deviation)*

Table 9. The results of the regression analysis for the example provided in Table 8.

Regression Statistics		
R Square	0,6946	
Adjusted R Square	0,4656	
Observations	8	
	Coefficients	***P-value***
Intercept	-91,07	0,10
x_1	143,58	0,10
x_2	0,82	0,10
x_3	0,01	0,96

As it can be observed, even though R^2 is around 70%, which is generally acceptable for assessing the performance of a regression model, R^2_{adj} is lower (about 47%). Therefore, we can conclude that there is still a significant portion of the variation in the dependent variable that is not explained by the model.

To improve the quality of the model, one possible solution could be to add

interaction terms in the model (in our case we choose x_1x_2 as these two factors are the most likely to influence the mass flow rate (Marushka et al., 2022)). The regression model will be as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 \quad (6)$$

Table 10. The updated 2^3 full factorial design matrix with the coded variables for the factors x_1 (bulk density), x_2 (orifice diameter), x_3 (relative humidity) with interaction x_1x_2, based on the example provided in Table 8.

Run	x^c_1	x^c_2	x^c_3	$x^c_1 x^c_2$
1	-1	-1	-1	1
2	1	-1	-1	-1
3	-1	1	-1	-1
4	1	1	-1	1
5	-1	-1	1	1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	1

Table 11. The natural values for the factors x_1 (bulk density), x_2 (orifice diameter), x_3 (relative humidity) with interaction x_1x_2 and the outcome mass flow rate (Y).

Run	x_1	x_2	x_3	x_1x_2	Y	SD
1	0,588	6	30	3,53	1,96	0,01
2	0,639	6	30	3,83	2,41	0,17
3	0,588	15	30	8,82	2,37	0,01
4	0,639	15	30	9,59	16,65	0,32
5	0,588	6	60	3,53	2,12	0,05
6	0,639	6	60	3,83	2,51	0,20
7	0,588	15	60	8,82	2,70	0,13
8	0,639	15	60	9,59	16,87	0,37

**replicated 10 times*

Table 9 displays the results of the regression analysis. The high correlation coefficients R^2 and R^2_{adj} indicate a perfect fit, all coefficients exhibit statistical significance.

Table 12. The results of the regression analysis for the example provided in Table 11.

Regression Statistics	
R Square	0,99995
Adjusted R Square	0,99989
Observations	8

	Coefficients	***P-value***
Intercept	102,67	6,9E-06
x_1	-172,22	6,3E-06
x_2	-17,63	9,1E-07
x_3	0,01	2,6E-02
x_1x_2	30,08	7,9E-07

The final model with the coefficients will be as follows:

$$Y = 102{,}67 - 172{,}22x_1 - 17{,}63x_2 + 0{,}01x_3 + 30{,}08x_1x_2$$

By substituting the values for the variables x_1, x_2 and x_3 one can calculate the predicted outcome $\hat{Y}_i$ and prediction precision (Δ (%)).

Table 13. The results of analysis for the example with interaction (Table 11) with the predicted value ($\hat{Y}_i$) of the mass flow rate, residual (e_i) and prediction precision (Δ).

Run	**Mass flow rate**	**Predicted Value *($\hat{Y}_i$)***	**Residual *(e_i)***	***Δ (%)***
1	1,96	2,05	0,09	4,59
2	2,41	2,47	0,06	2,49
3	2,37	2,56	0,19	8,02
4	16,65	16,79	0,14	0,84
5	2,12	2,35	0,23	10,85
6	2,51	2,77	0,26	10,36
7	2,70	2,86	0,16	5,93
8	16,87	17,09	0,22	1,30

3.1.2 FRACTIONAL FACTORIAL DESIGN

A fractional factorial design is a factorial design in which some of the possible combinations of factor levels (usually at least half) are skipped (Table 14). Fractional factorial designs offer the advantage of conserving resources by decreasing the size of the test sample, but this advantage comes with the cost of losing some information.

In a full factorial experiment, every factor has a designated experimental plan. In a fractional factorial experiment, a plan is established only for a subset of factors, which are referred to as the primary factors and mean the main factors of interest that are directly studied to understand their effects on the response variable. The plan for the remaining factors, termed secondary factors, depends on the primary factors, which helps to cut down the number of experimental trials. Secondary variables are often additional factors that researchers include in the experimental design to account for potential confounding effects. It is important to note that the terms "primary" and "secondary" do not imply any magnitude of their impact on the observed outcome *Y*. At the onset of the experiment, this impact cannot be predefined because the whole premise is based on the lack of initial knowledge about the factors' influence.

The number of experimental runs in case of a fractional factorial design for two levels is $n = 2^{k-p}$, where 2 means the number of factor's levels, k represents the total number of factors and 2^{-p} is the fraction of the full factorial design being implemented (i.e. if $p = 1$, half of the full FD is implemented, if $p = 2$, only a quarter, and etc). The general rule is that the number of runs n must not be less than the number of coefficients in the regression model (Tošenovský, 2010).

Table 14. Comparison of the number of runs for full factorial and fraction factorial designs with factors at two levels.

Number of factors	Experimental runs number full FD	Typical experimental run frac FD	Reduction
5	32	16	1/2
6	64	32	1/2
7	128	64	1/2
8	256	64	1/4
9	512	128	1/4
10	1024	128	1/8
11	2048	128	1/16
12	4096	256	1/16

For example, let us consider a 2^{3-1} design with x_1, x_2 and x_3 as factors. To obtain a one-half fraction (2^{3-1}) one should select two main factors (e.g., x_1, x_2), for which a full factorial design will be drawn up and secondary factor x_3 can be chosen such, that the corresponding factors satisfy $x_3 = x_1x_2$. Such a relationship of the factors is called a *generator* and for 2^{k-p} p indicates the number of secondary factors. By multiplying the generator by the left side, i.e. by the factor x_3, we will obtain the *generator equation* $I = x_1x_2x_3$, where I is the identity column (with all levels at 1) and where the number of letters represents the *resolution* of the fractional factorial design (for our case it is III and the plan is denoted 2_{III}^{3-1}).

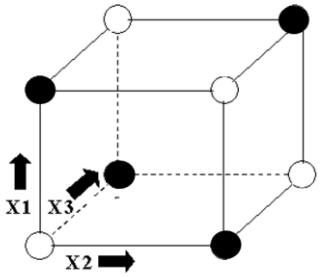
By definition, a full factorial design can be divided into two half fractions: a principal fraction (the fraction generated by positive design generators) and an alternate fraction. Which fraction to run is not critical unless one fraction is particularly difficult to run or one fraction contains a test combination of particular interest to the experimenters/subject matter experts. The information in either fraction is statistically equivalent. The typical design matrix for full factorial 2^3 and fractional factorial 2^{3-1} designs with visual representation is depicted in Table 15.

Table 15. 2^3 full factorial and 2^{3-1} fractional factorial design corresponding to the principal fraction (associated with plus in the last column) marked in grey with three factors (x_1, x_2 and x_3) at two levels (1 and -1).

Run	***I***	x^c_1	x^c_2	x^c_3	$x^c_1x^c_2$	$x^c_1x^c_2x^c_3$
1	1	1	-1	-1	-1	1
2	1	-1	1	-1	-1	1
3	1	-1	-1	1	1	1
4	1	1	1	1	1	1
5	1	1	1	-1	1	-1
6	1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1
8	1	-1	-1	-1	1	-1

Table 16. 2^{3-1} half-fraction factorial design (principal fraction = black dots).

Experimental run	x^c_1	x^c_2	$x^c_3 = x^c_1 x^c_2$
1	1	-1	-1
2	-1	1	-1
3	-1	-1	1
4	1	1	1



The key drawback of a fractional factorial design is that it introduces the possibility of "aliasing" or "confounding", where the effects of some factor combinations cannot be separated from others. For instance, in the previous example, it is impossible to distinguish the effect of factor x_3 from the combined effect of factors x_1 and x_2 because of the generator $x_3 = x_1x_2$. Therefore, in constructing a fractional factorial design, the design's so-called resolution should be considered. Examples of resolution:

Resolution III: Two-factor interactions are aliased with the main effects, but no main effects are aliased with any other main effect. For example, in a drug formulation study, a resolution III design may be used to investigate the effects of three factors: excipient type (A), drug concentration (B), and mixing time (C) (this design is denoted 2_{III}^{3-1}).

Resolution IV: No main effect is aliased with other main effects or with any two-factor interaction, but two-factor interaction are aliased with each other (2_{IV}^{4-1}). In a pharmaceutical manufacturing process optimization study, a resolution IV design could be used to analyse the effects of four factors: temperature (A), pressure (B), reaction time (C), and catalyst concentration (D).

Resolution V: No main effect or two-factor interaction is aliased with any other main effector two-factor interaction, but two factor integration is aliased with three-factor interaction (2_V^{5-1}). For instance, in a drug delivery system development study, a resolution V design might be employed to examine the effects of five factors: polymer type (A), drug loading (B), pH of the release medium (C), stirring rate (D), and temperature (E).

Designs with resolution levels below III are not typically useful, as level I designs consist of a single experimental run, and level II designs have mutually

confounded main effects. The most common types of designs have resolution levels III, IV, and V.

As an example of the application of a fraction factorial design, we can consider the same problem as earlier but let us add more factors that can potentially affect the outcome. In this case, a 2-level, 5-factor fractional factorial experimental design was implemented. A 1/2 fractional factorial design was selected to reduce the number of experiments from 32 to 16 (as suggested in Table 14). The variables included in this process were: the bulk density of the pharmaceutical powder (0,588 and 0,639 g/mL), orifice diameter (6 and 15 mm), relative humidity (30 and 60 %), the temperature of the environment (15 and 25C) and atmospheric pressure (750- and 760-mm Hg). In this case, we obtain resolution V or 2_V^{5-1} ($I = x_1x_2x_3x_4x_5$) for a one-half fractional factorial design.

Table 19 displays the finalized fractional factorial design. The dependent variable will be the mass flow rate as in the previous examples.

Table 17. The natural values of the levels for the factors x_1, x_2, x_3, x_4 and x_5.

Factor	**Notation**	**Low level (-1)**	**High level (1)**
Bulk density (g/mL)	x_1	0,588	0,639
Orifice diameter (mm)	x_2	6	15
Relative humidity (%)	x_3	30	60
Temperature (°C)	x_4	15	25
Atmospheric pressure (mm	x_5	750	760

Table 18. 2^5 full factorial design with coded levels (the principal fraction is marked in grey).

Run	x^c_1	x^c_2	x^c_3	x^c_4	x^c_5	Y
1	-1	-1	-1	-1	-1	1,99
2	1	-1	-1	-1	-1	2,45
3	-1	1	-1	-1	-1	2,39
4	1	1	-1	-1	-1	16,67
5	-1	-1	1	-1	-1	2,18
6	1	-1	1	-1	-1	2,57
7	-1	1	1	-1	-1	2,76
8	1	1	1	-1	-1	16,89
9	-1	-1	-1	1	-1	1,98
10	1	-1	-1	1	-1	2,49
11	-1	1	-1	1	-1	8,01
12	1	1	-1	1	-1	16,69
13	-1	-1	1	1	-1	2,19
14	1	-1	1	1	-1	2,58
15	-1	1	1	1	-1	2,77
16	1	1	1	1	-1	16,87
17	-1	-1	-1	-1	1	1,96
18	1	-1	-1	-1	1	2,41
19	-1	1	-1	-1	1	2,37
20	1	1	-1	-1	1	16,65
21	-1	-1	1	-1	1	2,12
22	1	-1	1	-1	1	2,51
23	-1	1	1	-1	1	2,71
24	1	1	1	-1	1	16,88
25	-1	-1	-1	1	1	1,96
26	1	-1	-1	1	1	2,41
27	-1	1	-1	1	1	2,37
28	1	1	-1	1	1	16,55
29	-1	-1	1	1	1	2,13
30	1	-1	1	1	1	2,55
31	-1	1	1	1	1	2,73
32	1	1	1	1	1	16,87

Table 19. The half-fractional factorial experimental design extracted from the Table 18 for determination of the main effects. Experimental order was randomized.

Run	x^c_1	x^c_2	x^c_3	x^c_4	$x^c_5 = x^c_1 x^c_2 x^c_3 x^c_4$	*Y*
1	-1	-1	-1	-1	1	1,96
2	1	-1	-1	-1	-1	2,45
3	-1	1	-1	-1	-1	2,39
4	1	1	-1	-1	1	16,65
5	-1	-1	1	-1	-1	2,18
6	1	-1	1	-1	1	2,51
7	-1	1	1	-1	1	2,71
8	1	1	1	-1	-1	16,89
9	-1	-1	-1	1	-1	1,98
10	1	-1	-1	1	1	2,41
11	-1	1	-1	1	1	2,37
12	1	1	-1	1	-1	16,69
13	-1	-1	1	1	1	2,13
14	1	-1	1	1	-1	2,58
15	-1	1	1	1	-1	2,77
16	1	1	1	1	1	16,87

The main effects of the factors on mass flow rate are presented in Table 20.

Table 20. The results of regression for the mass flow rate using the example of the half-fractional experimental design represented in Table 19.

Regression Statistics	
R Square	0,6949
Adjusted R^2	0,5423
Observations	16

	Coefficients	***P-value***
Intercept	5,9713	0,0003
x_1	3,6600	**0,0073**
x_2	3,6963	**0,0069**
x_3	0,1088	0,9225
x_4	0,0038	0,9973
x_5	-0,0200	0,9857

Examination of the main influences reveals that the coefficients for the factors x_1 (bulk density) and x_2 (orifice diameter) are statistically significant (*p*-value is below 5%) and these factors have a real effect on the mass flow rate. No significant impact of the other factors on the mass flow rate was observed. We could have included

interaction terms, as they are not confounded with the main effects, but the models with interactions give the *p*-values higher than 5%.

There is a range of factorial designs beyond full FD and frac FD, each offering varying degrees of insight. The depth of understanding these designs provide depends on their resolution level, because it indicates how the main effects, and their interactions are confounded.

Choosing the right design is essentially figuring out the best way to sample the realm of possibilities. A broad variety of factorial designs exist, some of which are used to filter out crucial variables (Level III resolution), while others are used to characterize processes (Levels IV-V resolution) or optimize them (Level IV resolution). Certain designs, such as definitive screening or designs associated with response surface methodology (RSM), are derived from factorial designs and can be thought of as partial factorial designs that include points or runs that are not covered by standard factorial designs (Jankovic et al., 2021).

Table 21. Properties of full FD and frac FD.

Type	Model	Recommended number of factors	Levels
frac FD	Linear and interaction	from 3 to 6	2 or 3
full FD	Linear and interaction	up to 6	2 or 3

Table 22. Literature examples of using full FD and frac FD in pharmaceutical technology.

Dosage form	DoE	Application	Study
Dispersible tablets	full FD	Implementation of QbD approach in formulation development.	(Charoo et al., 2012)
Emulsion	full FD, D-optimal Design	Optimize the emulsion for electro spinning.	(Badawi et al., 2014)
Immediate release tablet	frac FD	Examining the relative impact of active pharmaceutical ingredient (API) properties, processing methods, and excipients variability on drug product quality attributes.	(Kushner et al., 2014)
Drug loaded microsponge incorporated in gel base	full FD	To optimize the formulation.	(Kumar et al., 2017)

Nanoparticles	full FD	Optimization, development, and validation of HPLC method for estimation of valsartan in nanoparticles.	(Kumar et al., 2015)
Polymeric nanosuspension	full FD	Application of QbD approach to study the effect of CMAs and CPPs on critical quality attributes (CQAs) and to improve the quality and safety of formulation.	(Srinivas et al., 2017)
Extended-release tablets	full FD	Preparation and scale-up of extended-release low dose tablets of bromopride.	(Ferreira et al., 2014)
ODT	full FD	Utilizing the design of experiment approach to formulate, evaluate and co-crystal of piroxicam.	(Panzade et al., 2017)
Solid Lipid Nanoparticles	full FD	Utilizing DoE approach to investigate the influence of pre-freezing conditions on the powder respirability.	(Maretti et al., 2016)
Excipients Micronization	full FD	Effect of grinding pressure, injector pressure and feed rate on the particulate attributes of micronized powders procured from the different size grades.	(Chavez et al., 2015)
Oral Drug Suspension	full FD	Statistical optimization of extraction process for quantification of valsartan in rabbit plasma.	(Srenivasa et al., 2017)
Film coated tablets	full FD	To assess formulation ruggedness and optimize composition of excipients.	(Badawy et al., 2016)
Emulsions	frac FD	Preparation of the formulations based on the fractional factorial design	(Vasiljevic et al., 2017)

3.2 OTHER TYPES OF DOE USED IN PHARMACEUTICAL TECHNOLOGY

3.2.1 SCREENING DESIGNS

Screening is a technique that selects the factors having a significant impact on the response. This technique improves the understanding of the system, making it possible to only retain the factors which have a real impact. Although there are many different techniques, the most popular in pharmaceutical technology are low-resolution designs, such as fractional factorial design (frac FD), Taguchi design (TD), and Plackett-Burman design (PBD). These designs require a low number of experimental runs, allowing for reduced time and resource expenditures. More information on each of these screening experimental designs is provided below (Beg et al., 2019).

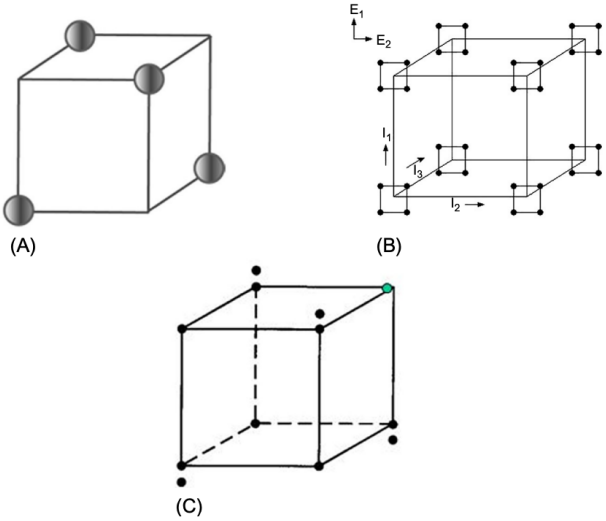


Figure 5. Examples of screening designs (A) fractional factorial design, (B) Taguchi design, and (C) Plackett-Burman design (Beg et al., 2019).

Fractional Factorial Designs (frac FD). This technique reduces the number of points per design (Figure 5A) as compared to full factorial designs and was treated in detail in subsection 3.2.

Taguchi's Designs (TD) were developed for industrial applications. These designs contain factorial designs which consider the interactions deemed important and reject any others. The aim of this technique is to make a product or process less variable (more robust) in the face of variation over which one has little or no

control. According to this method there are three types of factors involved in the process: signal factors (process control – e.g.: equipment settings), control factors (designer-controlled – e.g.: choice of materials), and noise factors (random variation – e.g.: operator skills). The method has three steps: concept design (targeting the final product), parameter design (identifying control factors for desired quality), and tolerance design (optimizing product performance).

An example of this design is depicted in Figure 5B, where I_1, I_2 and I_3 are factors that are under control or inner array factors (e.g.: composition of tablet, type of raw material, compression force) and signal outer array factors over which one has control only in the laboratory (temperature and relative humidity E_1 and E_2 accordingly). Fractional factorials could be used instead of full factorials for either the inner or outer array designs or for both. An example of the design matrix using the factors from the previous sections (namely: bulk density, orifice diameter and particle size as I_1, I_2 and I_3 and temperature and relative humidity as E_1 and E_2, response is the mass flow rate) is represented in Table *23*.

Table 23. The Taguchi's design matrix with the coded values for the factors I_1, I_2, I_3, E_1, E_2 with the outcome (mass flow rate).

Run				E_1	-1	1	-1	1	**Output**	**Output**
	I_1	I_2	I_3	E_2	-1	-1	1	1	**mean**	**SD**
1	-1	-1	-1		1,96	1,95	1,97	1,99	1,97	0,03
2	1	-1	-1		2,41	2,40	2,43	2,44	2,42	0,16
3	-1	1	-1		2,37	2,36	2,38	2,37	2,37	0,02
4	1	1	-1		16,65	16,64	16,73	16,70	16,68	0,29
5	-1	-1	1		2,12	2,10	2,15	2,13	2,13	0,05
6	1	-1	1		2,51	2,52	2,54	2,51	2,52	0,20
7	-1	1	1		2,70	2,61	2,71	2,78	2,70	0,13
8	1	1	1		16,87	16,86	16,88	16,80	16,85	0,36

The four outputs measured on each row correspond to the four outer array design points at each corner of the outer array box (see Figure 5B). As there are eight corners of the outer array box, there are eight rows in all. Each row yields a mean and standard deviation of mass flow rate. The desirable combination of factors would be the row that had both the highest average mass flow rate and the lowest standard deviation (variability).

Plackett-Burman designs (PBD) are designs for which the main aim is to study as many factors as possible in a minimum number of trials and to identify those that need to be studied in further rounds of experimentation in which interaction can be more thoroughly assessed (Figure 5C).

Plackett-Burman designs are usually resolution 2_{III}^{k-p} designs. In a resolution III design, main effects are aliased with 2-way interactions, therefore this design should be used to study main effects when it can be assumed that two-way interactions are negligible. PBD give designs with 12, 16, 20, 24, etc. number of runs (the number of experimental runs or trials is a multiple of four).

For better understanding let us consider a 2^5 design, using the factors and levels from the previous section (see subsection 3.1.1). In this case we have five factors, x_1 through x_5, each with two levels, and each factor is defined by a 12-run design, 6 plus ones and 6 minus ones. Half of the observations are at the high level and half at the low-level, and if one takes any two columns they are orthogonal to each other (if one takes the product of any two of these and add them up, the sum of the products will be zero. Because these are orthogonal , the factors are uncorrelated and one can get clean information on all main effects. The main effects are not confounded as dictated by the orthogonality of those columns. To create this type of design, first, one would fill out the first column of the design table, this would be the column x_1. Then one can create the x_2 column by taking the one but last element for the first position and then slide everything down. The last row must always be filled with -1. This process can be repeated for each column of factors needed in the design.

Table 24. The PBD matrix with coded values for the factors x_1, x_2, x_3, x_4, x_5 and the outcome (mass flow rate). The full matrix of 2^5 design is represented in Table 18.

Run	x^c_1	x^c_2	x^c_3	x^c_4	x^c_5	Y
1	1	-1	1	-1	-1	2,51
2	1	1	-1	1	-1	16,69
3	-1	1	1	-1	1	2,76
4	1	-1	1	1	-1	2,58
5	1	1	-1	1	1	16,69
6	1	1	1	-1	1	16,89
7	-1	1	1	1	-1	2,77
8	-1	-1	1	1	1	2,19
9	-1	-1	-1	1	1	1,98
10	1	-1	-1	-1	1	2,45
11	-1	1	-1	-1	-1	2,39
12	-1	-1	-1	-1	-1	1,99

Table 25. The comparison of the results obtained from the full factorial design of experiments and the Plackett-Burman design of experiments.

	Full Factorial Design	**Plackett-Burman design**
Number of runs	32	12
***p*-value for factor x_1**	0,000017	0,0192
***p*-value for factor x_2**	0,000004	0,0181
***p*-value for factor x_3**	0,924071	0,4042
***p*-value for factor x_4**	0,792289	0,3571
***p*-value for factor x_5**	0,768676	0,3597

As we can see, both designs give the same results in relation to the statistical significance of the factors. Despite the slight difference, in this example the factor settings and the conclusion remain the same when using either design, with an important difference in the number of experiments needed to be conducted to achieve these results.

Summarizing, the above Plackett-Burman design can be advantageous for:

- screening processes;
- when neglection of interaction is acceptable;
- two-level multi-factor experiments;
- experiments with more than four factors (otherwise a full factorial design can be employed);
- detection of large main effects;

Table 26. Literature examples of using screening designs in drug development.

Dosage form	DoE	Application	Study
Oral Delivery System	PBD	Enhanced biopharmaceutical characteristics in terms of PSD, encapsulation efficiency, and drug loading capacity.	(Javed et al., 2018)
Nanoparticles	TB	Enhanced capacity for drug loading and overcoming resistance to cancer drugs.	(Dong et al., 2009)
Transdermal films	PBD	The optimized formulation that resulted presents uniform thickness, a relatively low level of moisture absorption, and highly satisfactory drug loading.	(Ahmed et al., 2015)
Pellets	Frac FD	Identification of factors affecting layering powder efficiency.	(Tomuta et al., 2004)
Complex amphotericin B liposomal formulation	Frac FD	Identified that curing temperature during microfluidization has been identified as the most significant critical process parameter.	(Liu et al., 2020)

3.2.2 RESPONSE OPTIMIZATION

To optimize the response influenced by several variables one can chose to employ response surface methodology (RSM). RSM is a collection of mathematical and statistical techniques for the modelling and analysis of responses using visualization.

In most RSM tasks the true relationship between the dependent and independent variables is unknown. Therefore, the first step for a scientist is to find a suitable approximation of the form of relationship between Y and the set of independent variables x_i. This approximate relationship in the first phase could be linear in some small region, as in the first order model (*1*) (see chapter 3.1). Then, if there is a curvature detected in the system a polynomial of higher degree must be applied, such as a full second-order model:

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{j=1}^{k} \sum_{i<j} \beta_{ij} x_i x_j + \varepsilon$$

It is important to note that expecting a polynomial model to accurately represent the true functional relationship across the entire range of independent variables is unlikely. However, within a relatively small region, these models tend to perform quite well.

To estimate the coefficients for the approximating polynomials, the method of least squares, as discussed in chapter 3.1, is usually employed. Following this estimation, response surface analysis is conducted using the fitted surface. The fitted surface is the graph of the model plotted with the estimated coefficient values. In case of more than two factors 3D contour plots are used. If this fitted surface reasonably approximates the true response function, then analysing the fitted surface will yield results that are roughly equivalent to analysing the actual system. The effectiveness of coefficient estimation is greatly enhanced when proper experimental designs are used to collect the data (Montgomery, 2017).

RSM is useful for many problems in pharmaceutical industry which can be divided into the next three categories:

1. Mapping a response surface over a particular region of interest. It is used to predict the response that results from modifying or adjusting the process parameters (temperature of drying, tablet press machine settings).

2. Optimization of the response. An RSM study can help to estimate the region with the maximal response, which helps to find the optimum levels or conditions that should be chosen.
3. Selection of operating conditions to achieve specifications or customer requirements. In most response surface problems, there are several responses that should be simultaneously considered (amount of excipients, cost of production).

The most classic and commonly used design in combination with RSM is ***the Central Composite Design*** or CCD (a more detailed explanation will be represented in the next subsection). The CCD is a very efficient design for fitting second-order models for which just two parameters should be specified: the number of center points and a distance α from the design center. In general CCD consists of a 2^k full factorial or a 2^{k-p} fractional factorial design of resolution V with say n_f factorial runs, 2^k axial or star runs and n_0 center runs, k is the number of factors. The number of runs for CCDs is calculated by using of the following formula when the factorial design is full:

$$n = 2^k + 2k + n_0$$

In practice when a full 2^k FD exhibits lack of fit, one can extend to CCD by adding axial and center runs to allow the quadratic terms to be incorporated into the model.

The Box-Behnken Design, introduced by Box and Behnken in 1960, offers a set of three-level designs to model response surfaces. The Box-Behnken design is a quadratic design and does not include factorial or fractional factorial components. In this design, the treatment combinations are placed at the central points and midpoints along the edges of the process space. Box-Behnken designs are known for their high efficiency, they are requiring few experimental runs, and exhibit either rotatability or near-rotatability properties (explained in the next subsection).

Other Designs. In the case of two variables, designs composed of equally spaced points arranged in a circular fashion to create regular polygons can be employed. These configurations are often referred to as "equiradial designs" since the design points are evenly spaced from the center (Myers et al., 2016).

3.2.3 CENTRAL COMPOSITE DESIGN

The first part of a central composite design consists of a full or fractional factorial design. Apart from that, the central points within the experimental domain as well as the "star" points outside this domain make it possible to assess the response surface curvature (see Figure 6) (Ait-Amir et al., 2015).

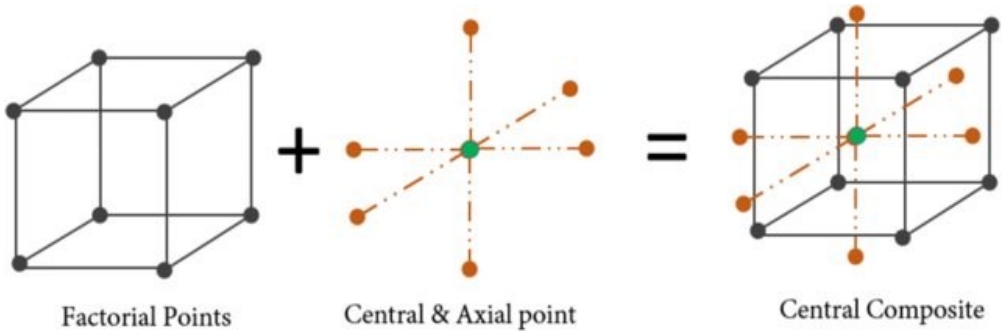


Figure 6. Generation of a central composite design (Boonpuang et al., 2020).

If the levels of the points of the factorial design part are ± 1, then for "star" points they are ± α where $|\alpha| \geq 1$. The selection of the value of α depends on the desired level of precision for estimating the surface response. The accuracy of this estimation is determined by the position of the star and centre points. Both the adjustment of the α value and the number of experimental runs at the centre of the domain influence the precision of the estimation (Ait-Amir et al., 2015).

Table 27. Short description and visualization of the three the most classic types of CCDs.

Type of the design	Description	Visualization
Circumscribed (CCC)	$\alpha = \pm\sqrt{2}$, the star points are located outside the experimental domain (distance α). The design requires five levels per factor.	CCC
Face-centered (CCF)	$\alpha = \pm 1$, the star points are located on the faces of the experimental domain. The design requires three levels per factor.	CCF
Inscribed (CCI)	$\alpha = \pm 1$, the design is used when it is not possible to leave the experimental domain. Used for situations in which the limits specified for variable settings are truly limited. CCI requires 5 levels per factor.	CCI

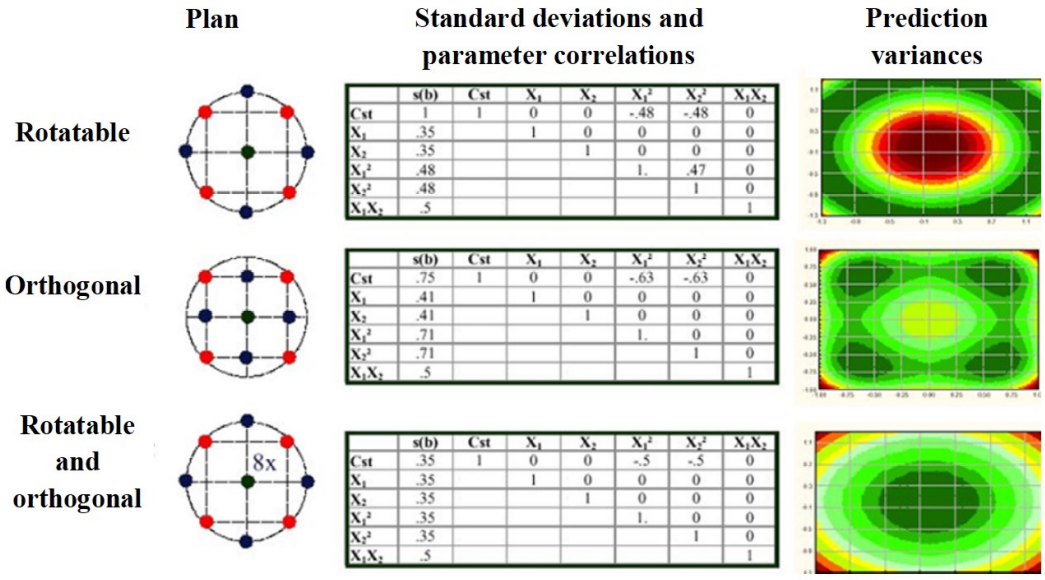
Two essential characteristics for CCD are orthogonality (see section 3.1.1) and rotatability. The design with the property to leave the prediction variance unchanged when the design is rotated around the center (0, 0, …, 0) is a rotatable design.

Based on these two characteristics, one can divide CCDs into three types:

Design	α	Number of centre points
Rotatable	$\alpha = \sqrt[4]{n_f}$	$n_0 = 1$
Orthogonal	$\alpha = \sqrt{\frac{\sqrt{n_f \cdot n} - n_f}{2}}$	$n_0 = 1$
Rotatable and orthogonal	$\alpha = \sqrt[4]{n_f}$	$n_0 = 4\sqrt{n_f + 4 - 2k}$

**parameters of various CC designs (where n_f is the number of trials of the factorial design part of experiments, n is the total number of trials of the experiment and k is the number of factors)*

A comparison of the quality of prediction for these three types of CC designs with two factors is represented below (Ait-Amir et al., 2015):



	s(b)	Cst	X_1	X_2	X_1^2	X_2^2	X_1X_2
Cst	1	1	0	0	-.48	-.48	0
X_1	.35		1	0	0	0	0
X_2	.35			1	0	0	0
X_1^2	.48				1.	.47	0
X_2^2	.48					1	0
X_1X_2	.5						1

	s(b)	Cst	X_1	X_2	X_1^2	X_2^2	X_1X_2
Cst	.75	1	0	0	-.63	-.63	0
X_1	.41		1	0	0	0	0
X_2	.41			1	0	0	0
X_1^2	.71				1.	0	0
X_2^2	.71					1	0
X_1X_2	.5						1

	s(b)	Cst	X_1	X_2	X_1^2	X_2^2	X_1X_2
Cst	.35	1	0	0	-.5	-.5	0
X_1	.35		1	0	0	0	0
X_2	.35			1	0	0	0
X_1^2	.35				1.	0	0
X_2^2	.35					1	0
X_1X_2	.5						1

The prediction variance (see section 3.1) changing from green to red means that the rotatable design gives less precise estimation than the orthogonal design at the centre of the experimental domain.

For illustration let us use the same variables from the previous subsections for determination of a mathematical model for the mass flow rate prediction. Let us consider two variables x_1 and x_2 with five levels (see Table 28).

Table 28. The natural and coded values for the variables x_1 (geometric mean of the particle size fraction or distribution) and x_2 (orifice diameter rounded to mm).

Level of factor	Orifice diameter (mm)	Particle size distribution (mm)
-α (star point)	3	0,059
-1 (low)	5	0,100
0 (centre point)	10	0,200
1 (high)	15	0,300
α (star point)	17	0,341

For determination of the uncoded value of alpha, the following equation can be used:

Uncoded value = coded value * (distance in real units between centre point and level 1) + centre point value in real units (see also (6)).

Table 29. The CCD matrix with the coded and natural values for the factors x_1 (orifice diameter) and x_2 (particle size) that allows rotatability and orthogonality and the outcome (mass flow rate).

Run	x_1	x_2	x^c_1	x^c_2	Mass flow rate (g/s)
1	5	0,100	-1	-1	1,91
2	15	0,100	1	-1	19,78
3	5	0,300	-1	1	2,4
4	15	0,300	1	1	25,5
5	3	0,200	$-\sqrt{2}$ (-α)	0	1,25
6	17	0,200	$\sqrt{2}$ (α)	0	28,45
7	10	0,060	0	$-\sqrt{2}$ (-α)	7,92
8	10	0,340	0	$\sqrt{2}$ (α)	10,39
9	10	0,200	0	0	10,28

Table 30. The results of analysis for the example in Table 28 for a simple first-order model.

Regression Statistics		
R Square	0,9476	
Adjusted R Square	0,9301	
Observations	9	
	Coefficients	***P-value***
Intercept	-10,42	0,01109
x_1	2,00	**0,00005**
x_2	12,21	0,25303

As we can observe the first prediction model has a high accuracy (R^2_{adj} – 93%) and will be as follows:

$$Y = 2x_1 + 12{,}21x_2 - 10{,}42$$

the *p*-value for the orifice diameter being lower than 5%.

Table 31. The results of analysis for the data provided in Table 29 for a second-order model (CCD).

Regression Statistics		
R Square	0,9980	
Adjusted R Square	0,9946	
Observations	9	
	Coefficients	***P-value***
Intercept	0,92	0,7826
x_1	-0,47	0,3142
x_1^2	0,10	**0,0121**
x_2	5,05	0,8126
x_2^2	-47,49	0,3649
Interaction	2,62	**0,0401**

The formula in this case will be as follows:

$$Y = 0{,}92 + 0{,}1x_1^2 - 0{,}47\,x_1 - 47{,}49\,x_2^2 + 5{,}05\,x_2 + 2{,}62\,x_1\,x_2$$

As it can be seen, the second-order model has the higher value of R^2_{adj} – 99,5%, which gives us more understanding of the data variability comparing to the first simple model. The CCD model allows to detect the statistical significance of interaction.

CCD was employed for the study of planetary ball milling of the pharmaceutical excipients for use in dosage forms with poorly soluble drugs.

The results were published as a first-author publication (Marushka et al., 2022):

MARUSHKA J.*, BROKEŠOVÁ J.*, OGADAH C.U., KAZEMI A., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: Milling of pharmaceutical powder carrier excipients: Application of central composite design. *Advanced Powder Technology,* 2022, 33(12), 103881. DOI:10.1016/j.apt.2022.103881, ISSN: 0921-8831, IF_{2022} 5,.2, Q_{AIS} 2

These results were also presented at the following international conference:

MARUSHKA J., BROKEŠOVÁ J., KAZEMI A., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: Design of experiments (DoE) to optimize the milling process of pharmaceutical powders. 13. Central European Symposium on Pharmaceutical Technology (CESPT), Gdansk, Poland, 16. - 18. 09. 2021

A model based on CCD was also used for the project No. 70119 of Grant Agency of Charles University. OGADAH C. U., MARUSHKA, J., ŠKLUBALOVÁ, Z, VRANÍKOVÁ, B. *Development of colon-targeted liquisolid systems for the local therapy of inflammatory bowel diseases*.

Table 32. Literature examples of using CCD in drug development.

Dosage form	DoE	Application	Study
Mixtures	CCD	Optimization of the mixture content	(Li et al., 2021)
Mixture of API polymer	CCD	Improvement of pharmaceutical characteristics of telmisartan	(Ritu et al., 2023)
Powder	CCD	Optimization of the planetary ball mill parameters	(Sharma et al., 2022)
PVA/CB composite	CCD	Prediction of surface roughness and dimensional accuracy in 3D printing	(Gregor et al., 2023)
Transdermal hydrogel	CCD	Optimization of ethosomal formulation	(Bhattacharya, 2021)
Nanoparticles	CCD	Optimization of solid lipid nanoparticles to enhance oral bioavailability	(Hassan et al., 2021)
Nanoparticles	CCD	Optimizing the biosynthesis of nanoparticles	(Nikaeen et al., 2020)

3.2.4 MIXTURE DESIGNS

In a pharmaceutical formulation, common applications encompass determining suitable diluent proportions in solid dosage forms, selecting optimal solvent-cosolvent combinations in liquid forms, and more application where mixture must be determined. The specificity of mixture is that the proportions of different components are interdependent. As such, an increase in one component's proportion necessitates a decrease in the proportion of one or more of the remaining components.

Mixture designs can take several forms, including simplex-lattice designs, simplex-centroid designs, axial designs, and extreme vertex designs. Moreover, extreme vertex designs consist of different types of so-called optimal designs, such as D-optimal, I-optimal, G-optimal, and A-optimal (they are described later in this section). These designs typically utilize three levels for each selected factor (Beg et al., 2019; Politis et al., 2017).

Simplex Lattice Design. In a mixture experiment, often polynomial functions without intercept are used to express the response, representing how the response is influenced by the components (Dejaegher et al., 2011). In a simplex design of degree m, each component has $m + 1$ distinct values, allowing the experimental results to fit a polynomial equation of a degree up to m. If q denotes the number of mixture components in the design, a $\{q, m\}$ simplex lattice design for q components comprises points defined by coordinate settings wherein each component assumes $m + 1$ equally spaced values ranging from 0 to 1 (see Figure 7).

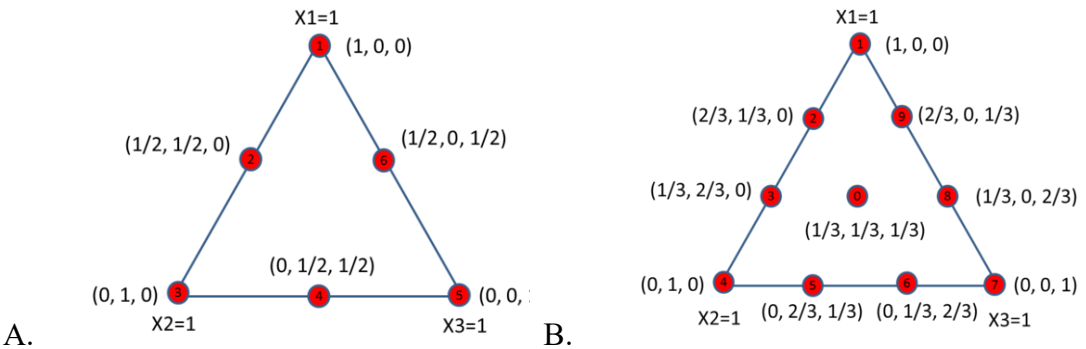


Figure 7. Typical simplex lattice with the point distribution for {3, 2} (A) and {3, 3} (B) design[3].

[3] https://help.reliasoft.com/reference/experiment_design_and_analysis/doe/mixture_design.html

In a simplex design of degree m, each component has $m + 1$ distinct values, allowing the experimental results to fit a polynomial equation of an order up to m.

Simplex Centroid Design. A simplex centroid design additionally involves central points. A simplex centroid design is more precise than a simplex lattice design and ensures that the design points are not skewed towards any region of the experimental space. It provides equal representation of the response variable in the central region, which can be important for understanding the behaviour of the system or process under investigation.

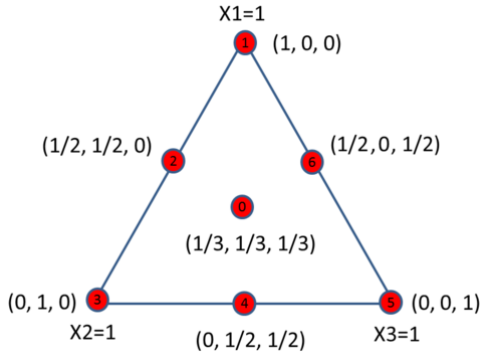


Figure 8. Simplex centroid design with the point distribution for 3 components[4].

For example, a {3, 2} simplex centroid design can be employed to fit the following model which is called the special cubic model:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_{12} X_1 X_2 + \beta_{13} X_1 X_3 + \beta_{23} X_2 X_3 + \beta_{123} X_1 X_2 X_3 \quad (6)$$

Simplex Axial Design. Simplex lattice and simplex centroid designs are classified as boundary designs because their points are located on the boundaries (such as vertices, edges, faces) of the simplex factor space, with the sole exception being the overall centroid. Conversely, axial designs primarily comprise points that are situated within the simplex. Axial designs are suggested for use when it is crucial to measure component effects in a screening experiment, especially when the goal is to fit first-degree models and when pure blends are not the focus (Dejaegher et al., 2011).

In a simplex axial design, all points are located on an axis. The axis of component i is a line from the base point (0) to the opposite vertex (1) represented in Figure 9. The most straightforward form of axial design is one where the points are evenly spaced from the overall centroid (the point with coordinates *1/q*, *1/q*, *1/q*...).

[4] https://help.reliasoft.com/reference/experiment_design_and_analysis/doe/mixture_design.html

Conventionally, the points that are halfway between the overall centroid and the vertex are referred to as axial check blends (here is it the points 4, 5, 6). A simplex axial design typically involves a first-degree model.

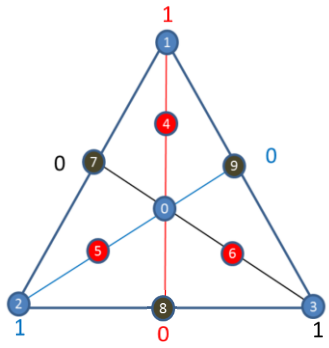


Figure 9. A simple axial design for 3 components[5].

Extreme Vertex Designs (Optimal Mixture Designs). Optimal designs use the "best" group of design points, selected from reducing or augmenting the number of experimental runs in the original design. Such designs are employed when components are subject to both lower and upper bound constraints (these designs cover only a subportion or smaller space within the simplex). Figure 10 depicts a typical optimal mixture design where the distribution of points fulfils the constraints for three components. In these designs, a pseudo-coding approach is adopted when both lower and upper bounds are in use. The user is required to determine the type of coding (pseudo, real, or actual) to be used for model fitting (Myers et al., 2016).

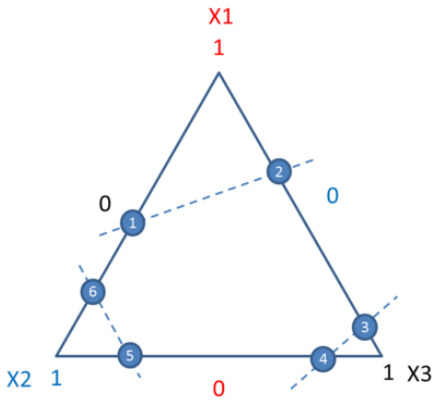


Figure 10. An extreme vertex design for 3 components[5].

[5] https://help.reliasoft.com/reference/experiment_design_and_analysis/doe/mixture_design.html

Optimal mixture designs are not restricted to mixture designs and can be further categorized into four types, each based on different optimality criteria: D-optimal, A-optimal, I-optimal, and G-optimal designs (Wu, 2011).

- D-Optimal Design. A D-optimal design minimizes the determinant of the Fisher information matrix (also known as the D-optimality criterion). The information matrix measures the precision of the parameter estimates in a regression model. By minimizing the determinant of the information matrix, a D-optimal design ensures the smallest possible confidence intervals for the estimated model coefficients that could be an advantage for factor screening which provides insights about critical factors.
- A-Optimal Design. This design aims to minimize the average variance of the polynomial coefficients (see subsection 3.1). It achieves this by seeking to minimize the sum of the diagonal elements of the information matrix. This criterion results in minimizing the average variance of the coefficient estimates based on a predetermined model.
- I-Optimal Design. Also known as IV-optimal (or integrated variance) design, this design minimizes the average prediction variance over the design space (I-optimality criterion). This criterion focuses on reducing the average uncertainty or variability in the predicted responses of the model across the range of the independent variables. An I-optimal design is useful when the primary interest is to accurately predict the response values.
- G-Optimal Design. A G-optimal design minimizes the maximum prediction variance over the design space (G-optimality criterion). This criterion aims to minimize the largest possible uncertainty or variability in the predicted responses of the model. A G-optimal design is particularly useful when the focus is on identifying the worst-case scenarios or detecting potential outliers in the response predictions (Beg, 2021; Jones et al., 2021).

Since D-optimal design is the most used in pharmaceutical sciences, as an example of application of this technique, let us consider an example of tablet mixture that contains lactose (filler or diluent), crospovidone (disintegrant), magnesium stearate (lubricant) and ibuprofen (API). One response, the mass flow rate (g/s), was examined. To analyse the results, a D-optimal mixture design was constructed, considering the

percentages of each component. The main factors with the lower and upper constraints are represented in Table 33.

Table 33. The main three factors of a D-optimal design at the upper and lower levels.

Ingredients	**Lower level (%)**	**Upper level (%)**
Lactose (*A*)	20	27
Crospovidone (*B*)	1	5
Magnesium stearate (*C*)	2	5

Points in the design space for D-optimal mixture design were prepared using Design-Expert Version 2023 software (Stat-Ease, Inc. 2023. Design-Expert®) and are indicated in the Table 34.

Table 34. The percentage of the components A, B, and C in 16 compositions of the tablet mixture, based on the D-optimal design, , along with the results of the mass flow rate measurements.

Mixture	***A***	***B***	***C***	***API***	***Mass flow rate (g/s)***
F1	20,00	5,00	5,00	70,00	1,98
F2	24,57	2,19	3,25	70,00	1,95
F3	24,00	1,00	5,00	70,00	2,41
F4	21,78	5,00	3,22	70,00	1,97
F5	24,00	1,00	5,00	70,00	2,32
F6	20,00	5,00	5,00	70,00	1,96
F7	24,41	3,59	2,00	70,00	2,11
F8	23,00	5,00	2,00	70,00	1,99
F9	21,40	4,10	4,51	70,00	2,18
F10	25,93	2,10	2,00	70,00	2,22
F11	26,00	1,00	3,00	70,00	2,45
F12	23,00	5,00	2,00	70,00	2,00
F13	25,90	2,10	2,00	70,00	2,39
F14	23,10	3,14	3,77	70,00	2,17
F15	26,00	1,00	3,00	70,00	2,49
F16	20,00	5,00	5,00	70,00	2,01

Based on the evaluation of the results, the best linear model that explains the influence of three factors on the mass flow is represented by the following linear equation:

$$Y = 2.91A - 1.21B - 14.50C \qquad (7)$$

Figure 11 illustrates visually the impact of the variables on the mass flow rate.

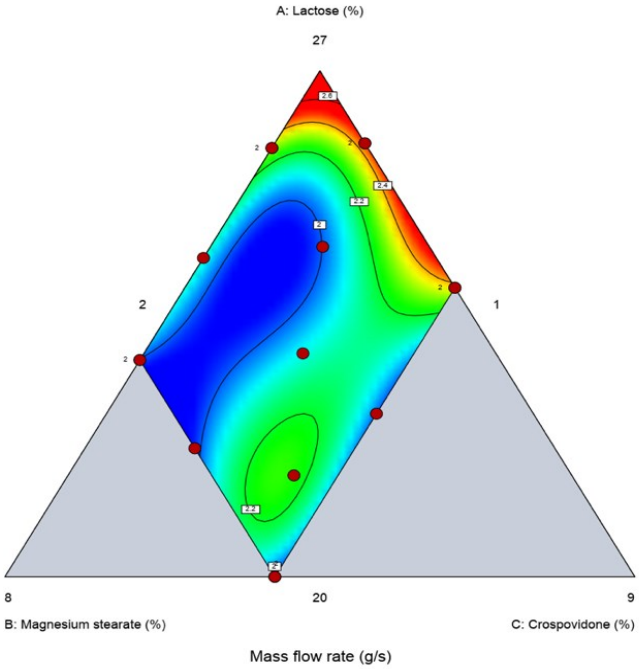


Figure 11. 2D contour plot representing the relationship between three variables: (A) the percentage of lactose, (B) the percentage of magnesium stearate, and (C) the percentage of crospovidone and their impact on the mass low rate.

As can be observed, a higher amount of disintegrant (component B) has a negative effect on the mass flow rate of the mixture (except for very small quantities – see red zones on the top of the plot). In contrast, expectedly a good result was shown by the addition of the lubricant (component A), which improved flow properties of the powder, as lubricants in general used to reduce friction between surfaces and improve the flow of materials, including powders. An interesting result was shown by a powder consisting only of pure lactose (without the addition of other excipients), which was also characterized by a good mass flow rate of the powder.

Table 35. A brief overview of properties of mixture designs.

Type	Model	Recommended number of factors	Levels
Simplex Lattice Design	Mixture model	from 2 to 4	3
Simplex Centroid Design	Mixture model	from 2 to 5	3
Constrained Mixture Design (Axial Design)	Mixture model	from 2 to 6	3
Extreme Vertex Design (Optimal Mixture design)	Mixture model	No limit	3

Table 36. Literature examples of using mixture design in drug development.

Dosage form	Application	Study
Gel	To estimate the effect of concentration of the components on the viscosity of the preparation.	(Cafaggi et al., n.d.)
Nanoparticles	Determination of drug-excipient compatibility.	(Pires et al., 2017)
Four-cosolvent blend	Verification of theophylline solubility in case of variations of the excipient mixture composition.	(Campisi et al., 1998)
Nanoparticles	Optimization D, L-lactic acid-based nanoparticles using D-optimal mixture design	(Adesina et al., 2014)
Tablets	Optimization of glibenclamide tablet composition using d-optimal mixture experimental design	(Mura et al., 2005)

3.3 OTHER DESIGNS AND DATA EXPLORATION METHODS

There are numerous situations in which classic experimental design techniques are no longer effective. The experimental design space may be constrained, or already-performed experiments may have to be included. The experiment may involve qualitative factors with more than two levels, mixture and process factors in the same design, or a specific set of design points. In addition, the situation may require reducing the number of experimental runs or using a reduced regression model in fitting the data. Finally, the region where the model is to be fitted may not be the same as where the measurements are to be made, or the model errors may have a known correlation matrix. In these cases, there are several other optimal designs which may be useful under different experimental situations.

For example, one can use computer-aided optimal designs that are particularly useful when classical designs do not apply (such as A, C, CD, CDT, D, DA, DS, DT, E, EA, G, I, KL, L, MS, S, T, U, V, and VS optimal design (Abd El-Monsef et al., 2011; Yue et al., 2011)). In these designs, each objective of the design problem is expressed as a convex function of the expected Fisher Information matrix (Lindner et al., 2006; Jung et al., 2021) and the optimal design is found by minimizing this function globally.

Principal Component Analysis (PCA). In this subsection we would like to stop on one of the most useful methods of handling of data already obtained.

Principal component analysis or PCA is one of the statistical techniques used to reduce the dimensionality of a large dataset while retaining the most important information in the data. This method helps to identify patterns and relationships between variables in a dataset, and to express these relationships in a way that is easier to understand and analyse than in the original variables.

PCA works by identifying the underlying structure of the data through the creation of new variables, called principal components (PC), that capture the maximum amount of variance in the data. These principal components are calculated as linear combinations of the original features, which means that each PC has a coefficient for each original variable. These coefficients, also called loadings, indicate how much each original feature contributes to the PC. The loadings can be positive or negative, which means that the original feature can have a positive or negative correlation with

the PC. The magnitude of the loadings reflects the importance of the original feature for the PC. The higher the absolute value of the loading, the more important the feature is.

The first principal component accounts for the largest amount of variance in the data, the second principal component accounts for the second largest amount of variance, and so on. First, to calculate the PCs from an original data matrix $\boldsymbol{X}$ with n rows for the samples and m columns for the variables, it is necessary to eliminate data heterogeneity by standardization. This is done by centering the mean in zero and the variance in 1, for each column from $\boldsymbol{X}$. The equation used for this purpose is the following:

$$Z_{i,j} = \sigma_j^{-1/2}(X_{i,j} - \mu_j);\ i = 1, n,\ j = i, \ldots, m,$$

where $X_{i,j}$ is the value of outcome j for the ith sample, $Z_{i,j}$ is the standardization of $X_{i,j}$, and μ_j and σ_j are the mean value and the variance of the outcome j, respectively. Then the covariance matrix $Z^TZ/(n-1)$ should be calculated; the desired PCs are its eigenvectors, in descending order with respect to the eigenvalues.

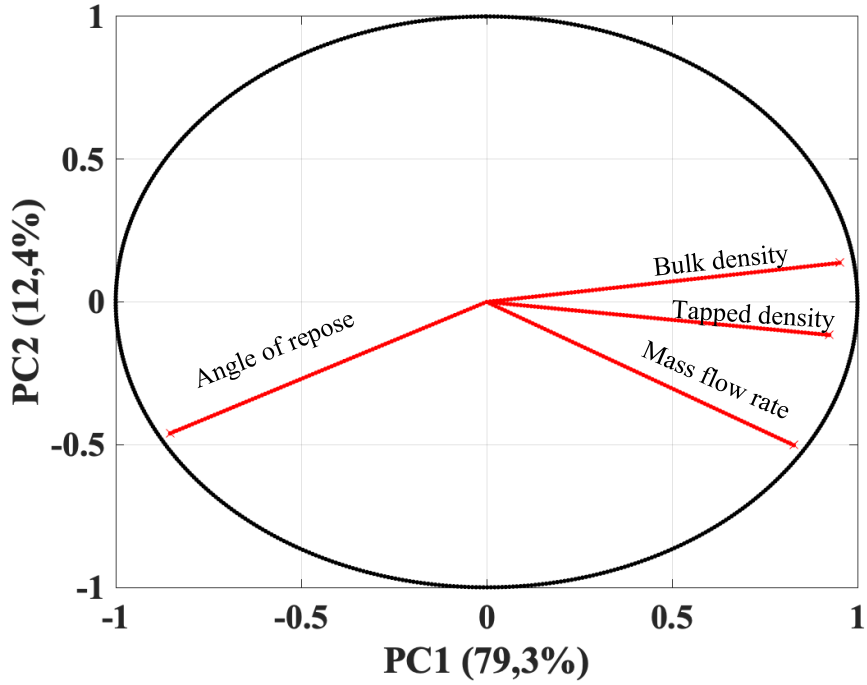
By reducing the dimensionality of the data using a few of these artificial principal components instead of all original outcomes, PCA can help to simplify complex datasets and highlight the most important patterns and relationships between variables. However, it is important to note that PCA assumes that the data is linearly related and normally distributed and may not be appropriate for all types of data.

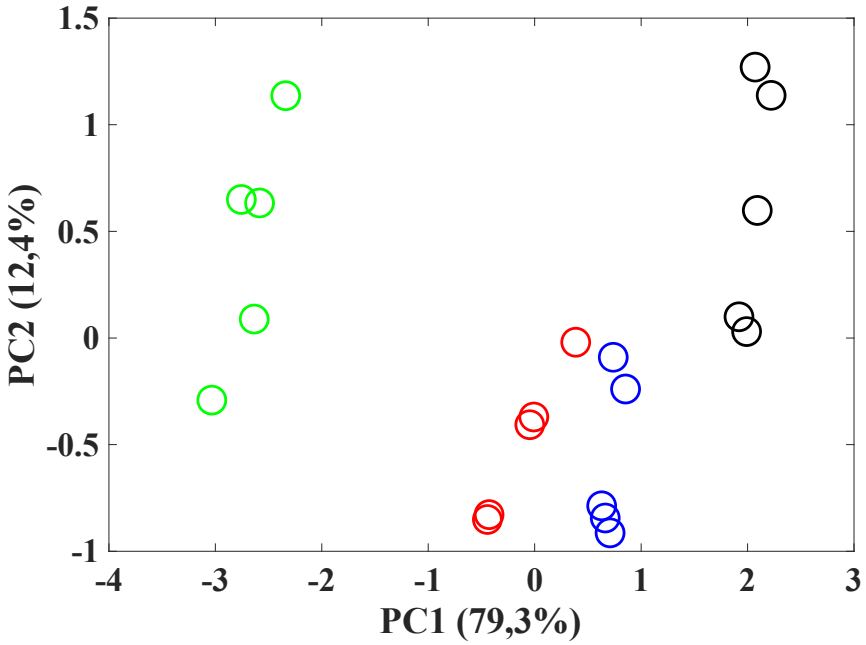
As an example of PCA application, let us use the same data from the previous chapters. In our case we will analyze four fractions of pharmaceutical powder with geometrical mean 0,100, 0,158, 0,245 and 0,345 mm. To obtain more data we will add a different amount of an excipient for lubrication of the particles (magnesium stearate (MgSt)), to each fraction in an amount from 1 to 5 %. As bulk density (d_b), tap density (d_t), angle of repose and mass flow rate are the main attributes that characterize bulk properties of powders we will present these variables as measured characteristics.

Table 37. The values for the independent variables and the data obtained.

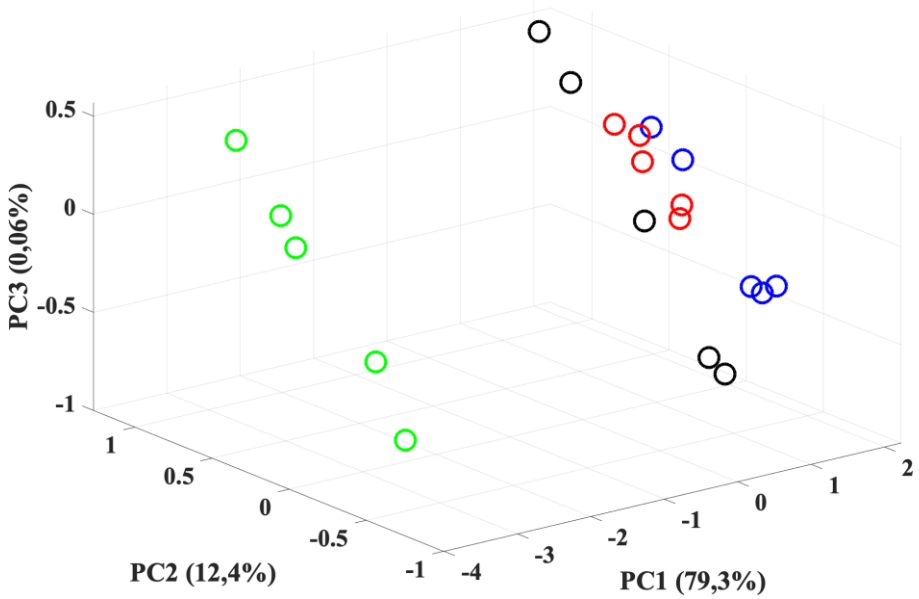
Powder fraction (mm)	MgSt content (%)	d_b (g/mL)	d_t (g/mL)	Angle of repose (°)	Mass flow rate (g/s)
0,100	1	0,590	0,650	40,00	1,98
	2	0,590	0,647	41,00	1,97
	3	0,587	0,644	39,00	1,95
	4	0,585	0,638	39,00	1,95
	5	0,588	0,640	38,00	1,94
0,158	1	0,613	0,662	39,00	2,42
	2	0,613	0,664	39,00	2,41
	3	0,616	0,665	38,00	2,41
	4	0,612	0,670	38,00	2,40
	5	0,610	0,680	37,00	2,39
0,245	1	0,611	0,700	38,00	2,49
	2	0,611	0,700	38,00	2,47
	3	0,612	0,698	38,00	2,46
	4	0,615	0,690	37,00	2,45
	5	0,618	0,682	37,00	2,43
0,345	1	0,638	0,724	37,00	2,38
	2	0,639	0,720	37,00	2,37
	3	0,640	0,711	36,00	2,37
	4	0,641	0,701	35,00	2,36
	5	0,644	0,690	35,00	2,35

A.





B.



C.

Figure 12. The loading plot (A) shows the correlation between the outcomes and the first and the second principal components. 2D and 3D PCA plots (B) and (C) show clustering of the observations depending on the particle size fraction (0,100 mm - green, 0,158 mm – red, 0,245 mm - blue, and 0,345 mm - black)[6].

[6]visualization was carried out using MATLAB, version 2023a (Natick, MA: The Math Works, Inc., 2023), accessed May 28, 2023, https://www.mathworks.com/

The loading plot shows the results for the first two components (see Figure 12A). It can be concluded that the outcomes mass flow rate, tapped density and bulk density are positively correlated with respect to each other and strongly positively correlated with PC1, which is responsible for almost 80% of variability in the data. Therefore, these three variables can be considered as the most sensitive outcomes. In its turn the angle of repose is negatively correlated with the other three parameters. Together PC1 and PC2 explain around 92% of the variance in the dataset.

The score plots (see Figure 12B and C) show that the first principal component separates the data into two clusters. The left cluster belongs to the smallest fraction (0,100 mm), while the right cluster is formed from the remaining three fractions. It confirms that particle size has a large influence on the variability of the outcome (Goh et al., 2018).

PCA was employed in two projects. The first work was related to the investigation of different coating material for the development of liquisolid systems, the results of calculation were included in the following publication (Vraníková et al., 2021):

VRANÍKOVA B., SVAČINOVA P., **MARUSHKA J.**, BROKEŠOVÁ J., HOLAS O., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: The importance of the coating material type and amount in the preparation of liquisolid systems based on magnesium aluminometasilicate carrier. *European Journal of Pharmaceutical Sciences,* 2021, 165, 105952. DOI:10.1016/j.ejps.2021.105952, ISSN: 0928-0987, IF_{2021} 5.112, Q_{AIS} 3

In the second project, in the pharmacology field, PCA was successfully used to analyse the factors influencing late pregnancy termination and their relation to the gene's expression. The results were presented in the following publication (Karahoda et al., 2021):

KARAHODA R., ROBLES M., ABAD C., **MARUSHKA J.,** STRANIK J., HORACKOVA H., DUINTJER TEBBENS J., VAILLANCOURT C., KACEROVSKY M., STAUD F.: Prenatal inflammation as a link between placental expression signature of tryptophan metabolism and preterm birth. *Human Molecular Genetics,* 2021, 30(22), 2053-2067. DOI:10.1093/hmg/ddab169, ISSN: 0964-6906, IF_{2021} 5.121, Q_{AIS} 1

4 RESULTS AND DISCUSSION

4.1 Flow Equations for Free-Flowable Particle Fractions of Sorbitol for Direct Compression: An Exploratory Multiple Regression Analysis of Particle and Orifice Size Influence.

MARUSHKA J., HURYCHOVÁ H., ŠKLUBALOVÁ Z., DUINTJER TEBBENS J.: Flow Equations for Free-Flowable Particle Fractions of Sorbitol for Direct Compression: An Exploratory Multiple Regression Analysis of Particle and Orifice Size Influence. *Pharmaceutics*, 2022, 14(8):1653.
DOI: 10.3390/pharmaceutics14081653. ISSN: 1999-4923, IF_{2022} 5.4, Q_{AIS} 2

The objective of this study was to investigate the relationship between the factors influencing the mass flow rate of a pharmaceutical powder. Based on the Beverloo equation, the bulk density, particle size and the orifice diameter play the main role in mass flow rate (Q_m) of a powder through a hopper orifice due to the gravity. In this research, we employ the Jones-Pilpel equation to model whether considering the interactions between these variables could notably enhance the precision of predicting the flow rate. In other words, we look for a mathematical model based on the Jones-Pilpel equation with logarithmical transformation to predict the mass flow rate (Q_m) of pharmaceutical powder with acceptably high precision of its prediction. To characterize the mass flow rate, as a model material, four different fractions (in the ranges of 0,080–0,125, 0,125–0,200, 0,200–0,300, and 0,300–0,400 mm) of the free flowable pharmaceutical excipient sorbitol were taken. After studying the properties of each fraction (scanning electron microscopy, bulk density, and mass flow rate), the data obtained were used to build eight mathematical models.

Since orifice diameter (D) and particle size (X) have the most pronounced effect on the mass flow rate, these two factors were considered as the main ones. Using four simple linear regression models for every particle size distribution separately, we achieved high precision of prediction with the average relative deviation of 1,27 – 5,27% between the experimentally measured and the predicted flow rate. If the simple model is used for the broader, entire particle size distribution, the precision of prediction decreases to approximately 14 %. Considering the influence of particle size

simultaneously through multilinear regression, the precision of prediction increases to 7.5 %; this can be further improved to 7 % when adding a term for interaction between orifice and particle diameter. A fully quadratic model (graphical representation is represented by Figure 13) achieves the high precision of the mass flow rate prediction of 3.1 %.

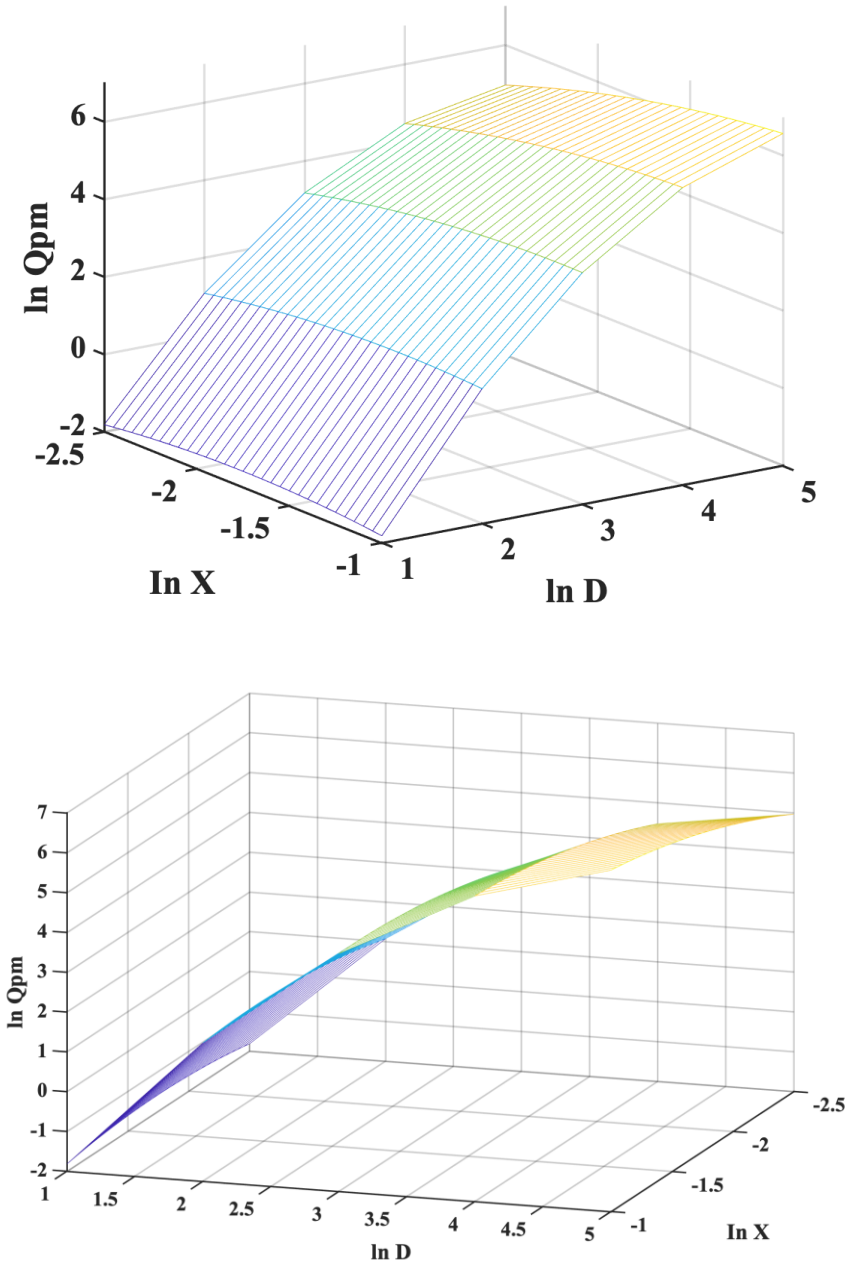


Figure 13. Quadratic regression 3D plot from two angles of view for predicted mass flow rate (logarithmized) in dependence of *ln D* and *ln X*.

Our primary findings indicate that (1) to achieve satisfactory prediction accuracy across a wide range of sorbitol particle sizes (0,1 to 0,346 mm), both orifice and particle diameter must be incorporated as factors in the regression analysis; (2) for optimal predictive performance, a fully quadratic model is necessary; (3) we have found a statistically significant interaction between orifice and particle diameter. As such materials as sorbitol are quite common in pharmacy, we expect that the outcomes of our study will be applicable to materials exhibiting similar granulometric characteristics.

4.2 Milling of pharmaceutical powder carrier excipients: Application of central composite design

MARUSHKA J.*, BROKEŠOVÁ J.*, OGADAH C.U., KAZEMI A., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: Milling of pharmaceutical powder carrier excipients: Application of central composite design. *Advanced Powder Technology,* 2022, 33(12), 103881.
DOI:10.1016/j.apt.2022.103881, ISSN: 0921-8831, IF_{2022} 5.2, Q_{AIS} 2

* *the authors contributed equally to this work*

In this article, we tried to investigate which of the most commonly used excipients in pharmaceutical technology are suitable for possible use in the preparation of binary interactive powder mixtures prepared by co-milling in a ball mill. The study was divided into two stages.

In the first part of the work, 24 types of excipients from different groups (lactose, celluloses, silicates, amino acids, alginates, starches, polyols, carrageenan and PVPs) were subjected to milling in a planetary ball mill under the same milling conditions (ball size 5 mm, time 15 min, milling speed 300 rpm). The most suitable criterion for evaluating the behavior of the material during milling is *span,* which is calculated based on the median size of the particles (x_{50}) and the width of the distribution (x_{10}, x_{90}): $span = (x_{90} - x_{10}) / x_{50}$. However, it is rare to see *span* as part of a particle size specification. The more common practice is to include two points which describe the coarsest and finest parts of the distribution (typically the x_{90} and x_{10}). In our investigation, it was decided to use both *span* and x_{90} as a characteristic criterion.

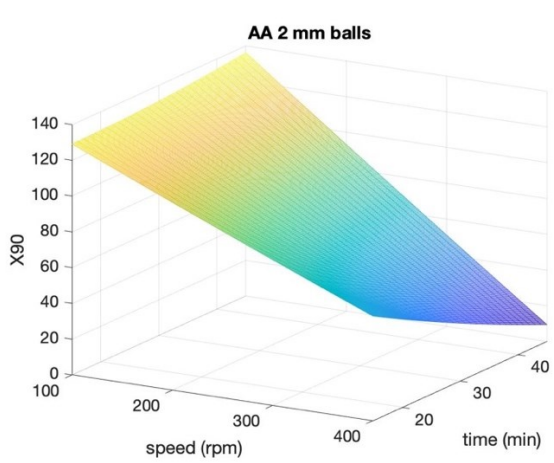
At the end of the first part of the experiments, only 10 out of 24 excipients with a x_{50} reduction of particles showed no signs of aggregation while maintaining a narrow *span* after the milling process. Substances that did not show suitable properties and their milling, in contrary, led to the agglomeration, hardening or sticking were excluded from the further experiment.

To evaluate the effect of the first factor (X_1), milling speed (100-400 rpm), and the second factor X_2, milling time (15-45 min), on the particle size and particle size distribution for three sizes of milling balls (2 mm, 5 mm, 10 mm) in the second part of the work, a central composite design (or CCD) was employed for selected substances. For this part only five excipients with the most favorable behavior were included in the investigation: alginic acid (AA), calcium alginate (CA), carrageenan (CAR), Avicel 200 (A200) and HPMC K15M (HPK). Using CCD, 30 experiments were planned and carried out to evaluate combinations of factors (independent variables) on the monitored responses *span* and x_{90}. For calculation of the prediction (y), the following quadratic model was used:

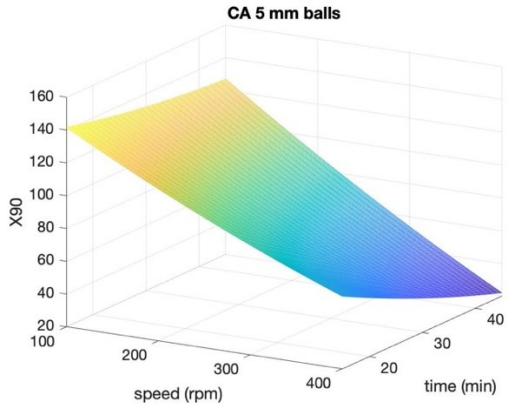
$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{11} X_1^2 + \beta_{22} X_2^2 + \beta_{12} X_1 X_2 + \varepsilon$$

where y is the expected value for the response, X_1 and X_2 are as mentioned above and β denotes regression coefficients characterizing the intercept (β_0), the main (β_1, β_2), quadratic (β_{11}, β_{22}), and the interaction (β_{12}) effects. The last term, ε, represents the error of the model.

The obtained surface plots (see Figure 14) for AA, CA and CAR illustrate the expected negative correlation between independent factors and particle size (x_{90}), where, depending on the size of the milling balls, effective particle reduction occurred at higher speed and longer milling time (Figure 14A, B and C). For A200, the lowest x_{90} values were achieved at a lower milling speed and using the largest milling balls (Figure 14D). A different behavior compared to the other abovementioned four materials was observed for HPK (Figure 14E).

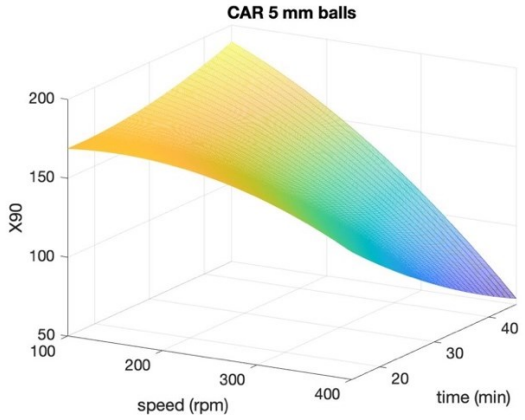


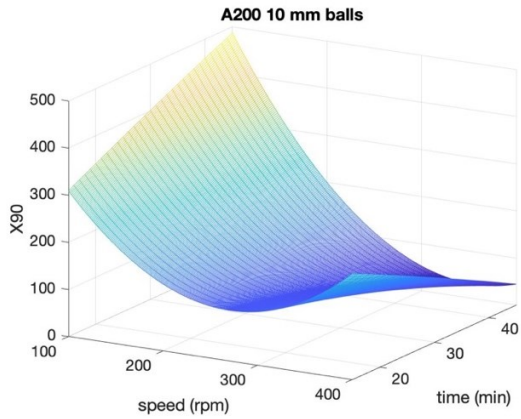
A.



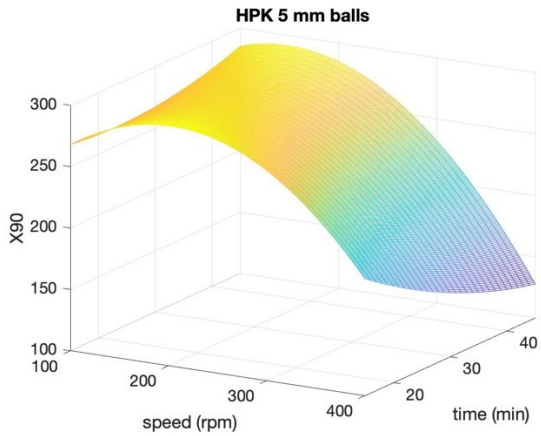
B.

C.





D.



E.

Figure 14. The response surface plots representing effects of milling speed (X_1) and milling time (X_2) on response (x_{90}) for (A) AA, (B) CA, (C) CAR, (D) A200 and (E) HPK at the optimal size of balls.

Finally, by utilizing the CCD, optimal milling conditions with respect to x_{90} (Table 38) were computed for the selected excipients AA, CA, CAR, A200, and HPK, considering their potential as carriers during milling with active substance. The experiment's results and comparison with the values predicted by the statistical model indicated that the predicted values were lower than the measured values, with the largest discrepancies observed for A200.

Table 38. CCD-determined optimal milling conditions and the x_{90} values achieved under the optimal conditions.

Excipient	**Milling ball size**	**Optimal milling conditions**	**Predicted value x_{90} (µm)**	**Observed value x_{90} (µm)**
AA	2 mm	400 rpm, 45	10,4	25,7
CA	5 mm	400 rpm, 45	22,8	36,9
CAR	5 mm	400 rpm, 45	55,2	96,5
A200	10 mm	356 rpm, 45	33,4	103,0
HPK	5 mm	400 rpm, 45	128,8	153,0

This discrepancy could be attributed to some limitation of CCD design such as a rather rigid data collection with five levels placed symmetrically on the scale around the central point as well as estimation of only the linear and quadratic terms and first order interaction. In addition, the results should be obtained under similar conditions. A varying number of balls of different size during milling according to CCD might influence stressing conditions due to a lower number of collisions (Descamps et al., 2016; Ju et al., 2019). The other reason of this discrepancy could be the influence of electrostatic forces generated during mixing and milling or the increased surface energy resulting from the greater surface area of smaller particles, which may cause particle agglomeration. Consequently, higher values might be detected during dry cell particle size measurements using laser diffraction. The thermal analysis did not reveal any changes in the crystalline or amorphous structure of the excipients after 45 minutes of milling in the screening experiment.

Although the particle size measured in real experiments was larger than the theoretical computations based on the CCD, the current study represents a valuable starting point in the application of CCD for milling processes of pharmaceutical powders. Further studies could include additional factors (for example described

above), levels or interactions that were not initially considered helping to improve the accuracy of predictions and model refinement, narrowing the gap between theoretical and real-world results.

4.3 The importance of the coating material type and amount in the preparation of liquisolid systems based on magnesium aluminometasilicate carrier.

VRANÍKOVA B., SVAČINOVA P., **MARUSHKA J.**, BROKEŠOVÁ J., HOLAS O., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: The importance of the coating material type and amount in the preparation of liquisolid systems based on magnesium aluminometasilicate carrier. *European Journal of Pharmaceutical Sciences,* 2021, 165, 105952.
DOI:10.1016/j.ejps.2021.105952, ISSN: 0928-0987, IF_{2021} 5.112, Q_{AIS} 3

While the formulation of liquisolid systems offers an innovative strategy for improving the dissolution of poorly soluble drugs, their widespread usage remains constrained primarily by challenges in transforming the liquid into a flowing and easily compressible powder. Therefore, the current study seeks to identify the optimal ratio (R value) of carrier to coating material for formulations utilizing magnesium aluminometasilicate (NUS2) loaded with polyethylene glycol 400.

In this study four types of commercially available colloidal silica (Aeroperl, Aerosil, Syloid 244 and Syloid 72) were employed as coating materials across nine different R values from 5 to 100 to evaluate the characteristics of powder and liquisolid compacts (flow rate and angle of repose of powder, angle of slide, compressibility index (CI), Hauser ratio (HR), energy of compression, uniformity of mass, pycnometric density and porosity, compact hardness, height and diameter and friability).

For investigation of the difference between coating materials properties the data analysis was conducted using principal component analysis (PCA), a technique that forms artificial combinations of all outcomes, called principal components (PCs), to analyse multivariate results. As the first three principal components accounted for 87.91% of the total variance in the original dataset, the discussion was focused on

these three components. In the PCA plots (Figure 15), it was evident that mixtures with the same coating material are closely situated and clustered into groups, indicating similar properties among these mixtures. Coating materials of similar nature, such as S244 and S72, are also grouped closely together. Moreover, scores plot (see (Figure 15A) says that the coating material does have an influence on the outcome variables. The biplot (Figure 15B), that is a loading plot and a scores plot merged, shows how each original variable contributes to the principal components.

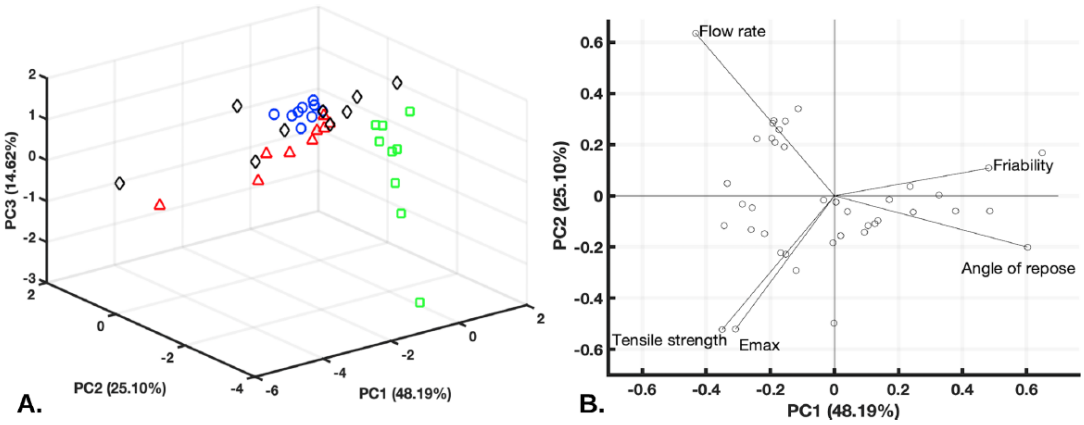


Figure 15. 3D PCA scores plot (A) demonstrates clustering of the observations depending on the type of coating material used (AS – green squares, AP – blue circles, S244 – red triangles, S72 – black diamonds). The PCA biplot (B) reflects the correlation between the variables and the first and second principal components (B).

Upon examining of the first PCs in the 2D biplot, it becomes clear that the results of measuring for angle of repose and friability are positively correlated with each other and have a strong positive correlation with PC1. As PC1 accounts for about half of the data variability (48%), these two characteristics were considered as the most sensitive among the five. The remaining outcomes are correlated with the second principal component, which explains roughly a quarter of the total variability. Notably, tensile strength and E_{max} have a strong correlation with each other.

These findings showed that the coating material does have an influence on the outcome. They are also underscoring the importance of considering PC1 and PC2 when interpreting the dataset, suggesting that the coating material does indeed influence the output. The sensitivity of the angle of repose and friability to PC1 indicates their significant role in affecting the overall variability. Also, the connection

between tensile strength and E_{max} points out a specific relationship between these two characteristics. This understanding of the principal components and their correlations enhances the comprehension of the interrelationships among the measured variables, providing a foundation for further exploration and interpretation of the dataset.

5 CONCLUSION

The primary aim of this study was to explore various experimental design and mathematical modeling techniques to address challenges in pharmaceutical technology. The selection of applied methods was based not only on the specific tasks at hand but also on the ease of implementing a particular experimental design technique.

The results of the three publications presented in this work lead to the following conclusions:

1. Our key findings suggest that effective mathematical modelling is crucial for studying powder characteristics, such as mass flow rate. We determined that incorporating both orifice and particle diameter as factors in regression analysis is essential for achieving satisfactory prediction precision across a broad range of sorbitol particle sizes (0,1 to 0,346 mm). For optimal predictive performance, a fully quadratic model is necessary, and the interaction between the factors is statistically significant. Focused on modelling the flow behaviour of the model excipient sorbitol for direct compression, the study clearly demonstrates the utility of mathematical analysis. Sorbitol, known for its free-flowing nature and relatively wide particle size distribution, represents common materials in pharmacy. Consequently, we expect that our study's outcomes will be applicable to materials sharing similar granulometric characteristics. The success of our mathematical modelling approach underscores its significance in enhancing the understanding and predictability of flow behaviour, particularly in the context of pharmaceutical materials.
2. Despite differences between real-world measurements and theoretical predictions in the study of the milling process of pharmaceutical powders using CCD, the CCD model suggests the region in the parameter space where the optimum may be found. To enhance prediction accuracy and refine the model, future studies could incorporate additional factors, levels, or interactions not initially considered. This iterative approach holds the potential to narrow the gap between theoretical expectations and real-world results, contributing to the ongoing improvement and applicability of CCD in pharmaceutical powder milling processes — an essential

technique for enhancing the solubility of poorly soluble drugs in pharmaceutical technology.

3. The use of PCA in modern research is crucial, allowing simultaneous study of a large amount of not only initial data but also several outcomes. In our study of liquisolid preparation, we demonstrated how to effectively analyse a large dataset and find connections between factors. These findings underscore the importance of considering PC1 and PC2 when interpreting the dataset. In our study the findings suggest that the coating material does influence the outcome. The sensitivity of the angle of repose and friability to PC1 underscores their significant influence on the overall variability within the dataset. Similarly, the observed correlation between tensile strength and Emax highlights a specific relationship between these two variables. This understanding of principal components and their correlations deepens our comprehension of how the measured variables interact, establishing a solid foundation for continued exploration and interpretation of the dataset and simplifying research in the field of liquisolid systems — an innovative strategy for improving the dissolution of poorly soluble drugs.

6 RESEARCH OUTPUTS

6.1 Articles related to the topic of the dissertation

MARUSHKA J., HURYCHOVÁ H., ŠKLUBALOVÁ Z., DUINTJER TEBBENS J.: Flow Equations for Free-Flowable Particle Fractions of Sorbitol for Direct Compression: An Exploratory Multiple Regression Analysis of Particle and Orifice Size Influence. *Pharmaceutics*, 2022, 14(8):1653.
DOI: 10.3390/pharmaceutics14081653. ISSN: 1999-4923, IF_{2022} 5.4, Q_{AIS} 2

Candidate's contribution:
First author, mathematical modelling, data analysis, interpretation of the results, visualization, writing the article, reviewing, and editing for submission

MARUSHKA J.*, BROKEŠOVÁ J.*, OGADAH C.U., KAZEMI A., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: Milling of pharmaceutical powder carrier excipients: Application of central composite design. *Advanced Powder Technology,* 2022, 33(12), 103881. DOI:10.1016/j.apt.2022.103881, ISSN: 0921-8831, IF_{2022} 5.2, Q_{AIS} 2

Candidate's contribution:
First author, development of experimental design, performing experiments (milling of the selected materials on the second phase of the study), data analysis, interpretation of the results, visualization, writing the article, reviewing, and editing for submission
** the authors contributed equally to this work*

VRANÍKOVÁ B., SVAČINOVÁ P., **MARUSHKA J.,** BROKEŠOVÁ J., HOLAS O., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: The importance of the coating material type and amount in the preparation of liquisolid systems based on magnesium aluminometasilicate carrier. *European Journal of Pharmaceutical Sciences,* 2021, 165, 105952. DOI:10.1016/j.ejps.2021.105952, ISSN: 0928-0987, IF_{2021} 5.112, Q_{AIS} 3

Candidate's contribution:
Data analysis, interpretation of the results, visualization, assisted in writing the article, reviewing, and editing for submission

6.2 Articles unrelated to the topic of the dissertation

KARAHODA R., ROBLES M., ABAD C., **MARUSHKA J.,** STRANIK J., HORACKOVA H., DUINTJER TEBBENS J., VAILLANCOURT C., KACEROVSKY M., STAUD F.: Prenatal inflammation as a link between placental expression signature of tryptophan metabolism and preterm birth. *Human Molecular Genetics,* 2021, 30(22), 2053-2067.
DOI:10.1093/hmg/ddab169, ISSN: 0964-6906, IF_{2021} 5.121, Q_{AIS} 1

Candidate's contribution:
Data analysis, interpretation of the results, visualization, assisted in writing the article, reviewing and editing for submission

6.3 Poster/oral presentations

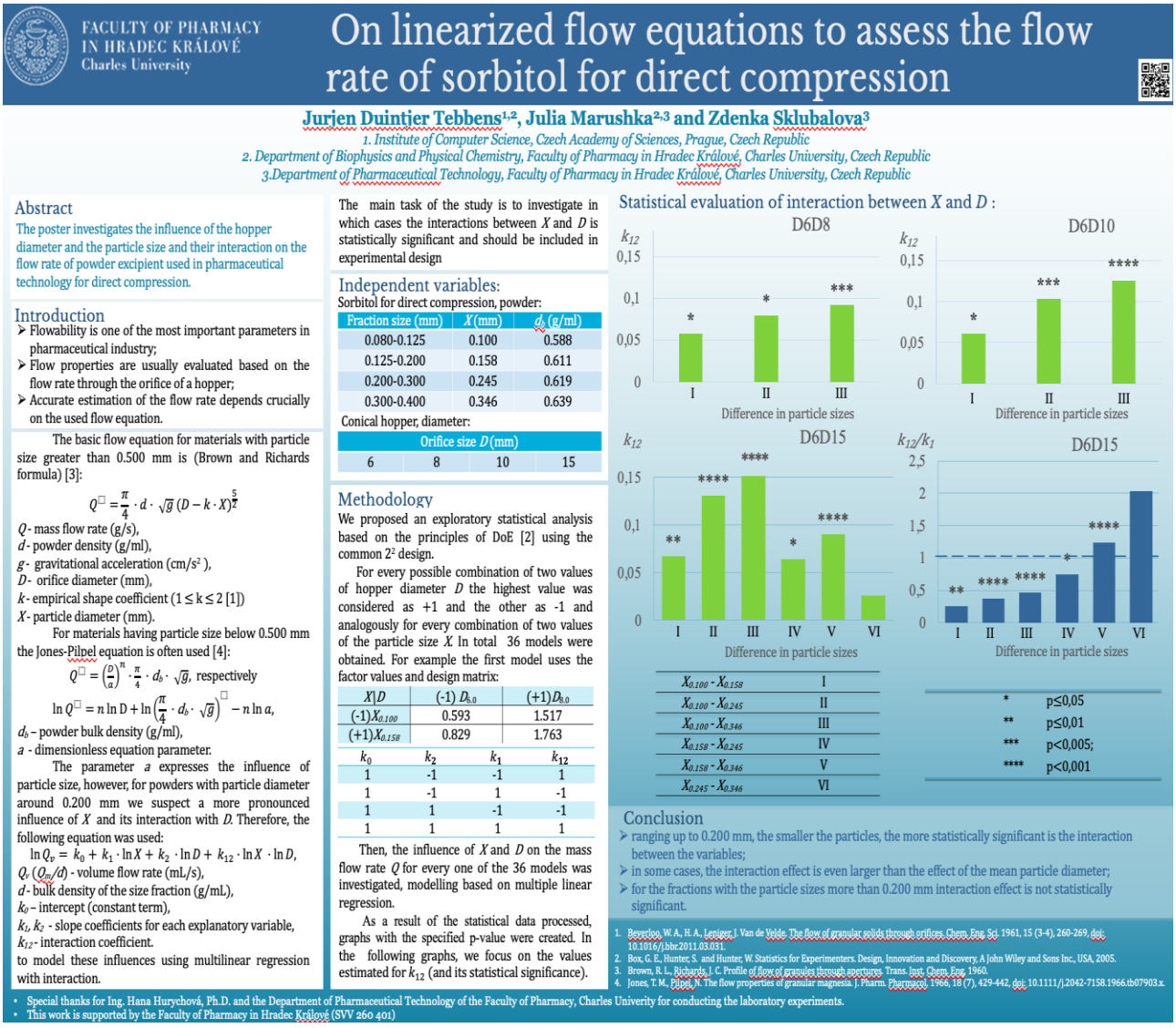
MARUSHKA, J., BLAGOVESTOVA KOVACHEVA M., RANTANEN, J., DUINTJER TEBBENS J., ŠKLUBALOVA, Z.: Study of the effect of water activity on the kinetics of theophylline monohydrate dehydration using a multivariate statistical approach. 12. Postgraduate and postdoctoral conference, Hradec Králové, CR, 01 - 02. 02. 2022

MARUSHKA J., BROKEŠOVA J., DUINTJER TEBBENS J., KAZEMI A., ŠKLUBALOVA Z.: Application of advanced design of experiments techniques for improvement of solubility of poorly soluble substances. 11. Postgraduate and postdoctoral conference, Hradec Králové, CR, 27 - 28. 01. 2021

MARUSHKA J., DUINTJER TEBBENS J., KARAHODA R., ŠTAUD F.: Introduction into the main ideas of PCA with illustrations from pharmaceutical technology and pharmacology. 10. Postgraduate and postdoctoral conference, Hradec Králové, CR, 22 - 23.01.2020

MARUSHKA J., DUINTJER TEBBENS J., ŠKLUBALOVA Z.: Determination of the significance of the influence of interaction between particle size and office diameter for the prediction of a flow rate of pharmaceutical filler for direct compression. X Congress of the Slovak Pharmaceutical Society, Bratislava, Slovakia, 05 – 07.09.2019

MARUSHKA J., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: On linearized flow equations to assess the flow rate of sorbitol for direct compression. 3rd IMA Conference on Dense Granular Flows - IMA, Cambridge, UK 01 - 04. 07. 2019



FACULTY OF PHARMACY IN HRADEC KRÁLOVÉ
Charles University

On linearized flow equations to assess the flow rate of sorbitol for direct compression

Jurjen Duintjer Tebbens[1,2], Julia Marushka[2,3] and Zdenka Sklubalova[3]

1. Institute of Computer Science, Czech Academy of Sciences, Prague, Czech Republic
2. Department of Biophysics and Physical Chemistry, Faculty of Pharmacy in Hradec Králové, Charles University, Czech Republic
3. Department of Pharmaceutical Technology, Faculty of Pharmacy in Hradec Králové, Charles University, Czech Republic

Abstract

The poster investigates the influence of the hopper diameter and the particle size and their interaction on the flow rate of powder excipient used in pharmaceutical technology for direct compression.

Introduction

- Flowability is one of the most important parameters in pharmaceutical industry;
- Flow properties are usually evaluated based on the flow rate through the orifice of a hopper;
- Accurate estimation of the flow rate depends crucially on the used flow equation.

The basic flow equation for materials with particle size greater than 0.500 mm is (Brown and Richards formula) [3]:

$$Q^{\square} = \frac{\pi}{4} \cdot d \cdot \sqrt{g}\,(D - k \cdot X)^{\frac{5}{2}}$$

Q - mass flow rate (g/s),
d - powder density (g/ml),
g - gravitational acceleration (cm/s^2),
D - orifice diameter (mm),
k - empirical shape coefficient ($1 \le k \le 2$ [1])
X - particle diameter (mm).

For materials having particle size below 0.500 mm the Jones-Pilpel equation is often used [4]:

$$Q^{\square} = \left(\frac{D}{a}\right)^{n} \cdot \frac{\pi}{4} \cdot d_b \cdot \sqrt{g}, \text{ respectively}$$

$$\ln Q^{\square} = n \ln D + \ln\left(\frac{\pi}{4} \cdot d_b \cdot \sqrt{g}\right)^{\square} - n \ln a,$$

d_b – powder bulk density (g/ml),
a - dimensionless equation parameter.

The parameter a expresses the influence of particle size, however, for powders with particle diameter around 0.200 mm we suspect a more pronounced influence of X and its interaction with D. Therefore, the following equation was used:

$$\ln Q_v = k_0 + k_1 \cdot \ln X + k_2 \cdot \ln D + k_{12} \cdot \ln X \cdot \ln D,$$

Q_v (Q_m/d) - volume flow rate (mL/s),
d - bulk density of the size fraction (g/mL),
k_0 – intercept (constant term),
k_1, k_2 - slope coefficients for each explanatory variable,
k_{12} - interaction coefficient.
to model these influences using multilinear regression with interaction.

The main task of the study is to investigate in which cases the interactions between X and D is statistically significant and should be included in experimental design

Independent variables:

Sorbitol for direct compression, powder:

Fraction size (mm)	X (mm)	d_b (g/ml)
0.080-0.125	0.100	0.588
0.125-0.200	0.158	0.611
0.200-0.300	0.245	0.619
0.300-0.400	0.346	0.639

Conical hopper, diameter:

Orifice size D (mm)			
6	8	10	15

Methodology

We proposed an exploratory statistical analysis based on the principles of DoE [2] using the common 2^2 design.

For every possible combination of two values of hopper diameter D the highest value was considered as +1 and the other as -1 and analogously for every combination of two values of the particle size X. In total 36 models were obtained. For example the first model uses the factor values and design matrix:

$X \backslash D$	(-1) $D_{8.0}$	(+1) $D_{8.0}$
(-1) $X_{0.100}$	0.593	1.517
(+1) $X_{0.158}$	0.829	1.763

k_0	k_2	k_1	k_{12}
1	-1	-1	1
1	-1	1	-1
1	1	-1	-1
1	1	1	1

Then, the influence of X and D on the mass flow rate Q for every one of the 36 models was investigated, modelling based on multiple linear regression.

As a result of the statistical data processed, graphs with the specified p-value were created. In the following graphs, we focus on the values estimated for k_{12} (and its statistical significance).

Statistical evaluation of interaction between X and D :

$X_{0.100}$ - $X_{0.158}$	I
$X_{0.100}$ - $X_{0.245}$	II
$X_{0.100}$ - $X_{0.346}$	III
$X_{0.158}$ - $X_{0.245}$	IV
$X_{0.158}$ - $X_{0.346}$	V
$X_{0.245}$ - $X_{0.346}$	VI

*	p≤0,05
**	p≤0,01
***	p<0,005;
****	p<0,001

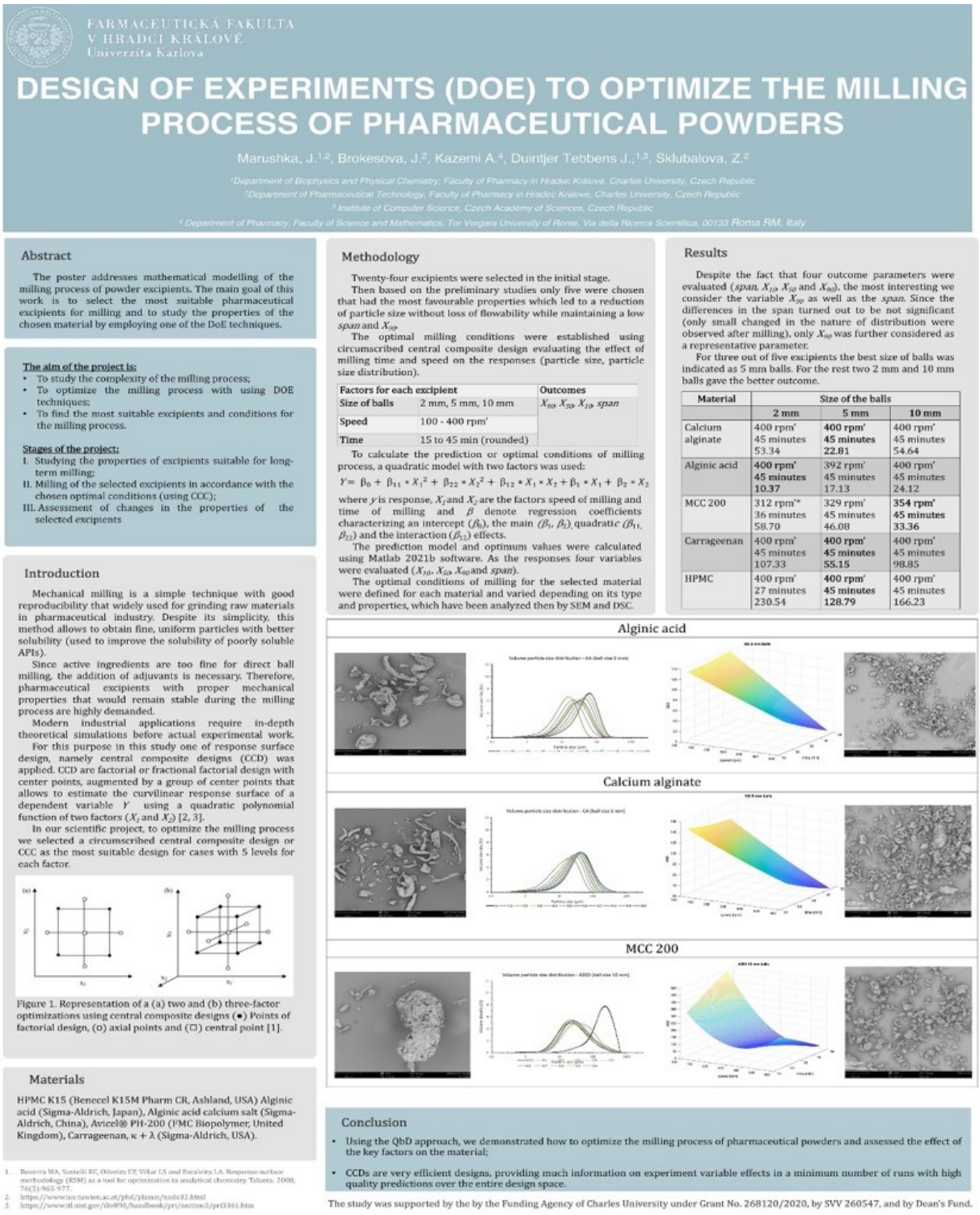
Conclusion

- ranging up to 0.200 mm, the smaller the particles, the more statistically significant is the interaction between the variables;
- in some cases, the interaction effect is even larger than the effect of the mean particle diameter;
- for the fractions with the particle sizes more than 0.200 mm interaction effect is not statistically significant.

1. Beverloo, W. A., H. A., Leniger, J. Van de Velde. The flow of granular solids through orifices. Chem. Eng. Sci. 1961, 15 (3-4), 260-269. doi: 10.1016/j.bbr.2011.03.031.
2. Box, G. E., Hunter, S. and Hunter, W. Statistics for Experimenters. Design, Innovation and Discovery, A John Wiley and Sons Inc., USA, 2005.
3. Brown, R. L., Richards, J. C. Profile of flow of granules through apertures. Trans. Inst. Chem. Eng. 1960.
4. Jones, T. M., Pilpel, N. The flow properties of granular magnesia. J. Pharm. Pharmacol. 1966, 18 (7), 429-442. doi: 10.1111/j.2042-7158.1966.tb07903.x.

- Special thanks for Ing. Hana Hurychová, Ph.D. and the Department of Pharmaceutical Technology of the Faculty of Pharmacy, Charles Univerity for conducting the laboratory experiments.
- This work is supported by the Faculty of Pharmacy in Hradec Králové (SVV 260 401)

MARUSHKA J., BROKEŠOVÁ J., KAZEMI A., DUINTJER TEBBENS J., ŠKLUBALOVÁ Z.: Design of experiments (DoE) to optimize the milling process of pharmaceutical powders. 13. Central European Symposium on Pharmaceutical Technology (CESPT), Gdansk, Poland, 16. - 18. 09. 2021



FARMACEUTICKÁ FAKULTA V HRADCI KRÁLOVÉ
Univerzita Karlova

DESIGN OF EXPERIMENTS (DOE) TO OPTIMIZE THE MILLING PROCESS OF PHARMACEUTICAL POWDERS

Marushka, J.[1,2], Brokesova, J.[2], Kazemi A.[4], Duintjer Tebbens J.,[1,3], Sklubalova, Z.[2]

[1]Department of Biophysics and Physical Chemistry, Faculty of Pharmacy in Hradec Králové, Charles University, Czech Republic
[2]Department of Pharmaceutical Technology, Faculty of Pharmacy in Hradec Králové, Charles University, Czech Republic
[3] Institute of Computer Science, Czech Academy of Sciences, Czech Republic
[4] Department of Pharmacy, Faculty of Science and Mathematics, Tor Vergata University of Rome, Via della Ricerca Scientifica, 00133 Roma RM, Italy

Abstract

The poster addresses mathematical modelling of the milling process of powder excipients. The main goal of this work is to select the most suitable pharmaceutical excipients for milling and to study the properties of the chosen material by employing one of the DoE techniques.

The aim of the project is:

- To study the complexity of the milling process;
- To optimize the milling process with using DOE techniques;
- To find the most suitable excipients and conditions for the milling process.

Stages of the project:

I. Studying the properties of excipients suitable for long-term milling;
II. Milling of the selected excipients in accordance with the chosen optimal conditions (using CCC);
III. Assessment of changes in the properties of the selected excipients

Introduction

Mechanical milling is a simple technique with good reproducibility that widely used for grinding raw materials in pharmaceutical industry. Despite its simplicity, this method allows to obtain fine, uniform particles with better solubility (used to improve the solubility of poorly soluble APIs).

Since active ingredients are too fine for direct ball milling, the addition of adjuvants is necessary. Therefore, pharmaceutical excipients with proper mechanical properties that would remain stable during the milling process are highly demanded.

Modern industrial applications require in-depth theoretical simulations before actual experimental work.

For this purpose in this study one of response surface design, namely central composite designs (CCD) was applied. CCD are factorial or fractional factorial design with center points, augmented by a group of center points that allows to estimate the curvilinear response surface of a dependent variable Y using a quadratic polynomial function of two factors (X_1 and X_2) [2, 3].

In our scientific project, to optimize the milling process we selected a circumscribed central composite design or CCC as the most suitable design for cases with 5 levels for each factor.

Figure 1. Representation of a (a) two and (b) three-factor optimizations using central composite designs (●) Points of factorial design, (o) axial points and (□) central point [1].

Materials

HPMC K15 (Benecel K15M Pharm CR, Ashland, USA) Alginic acid (Sigma-Aldrich, Japan), Alginic acid calcium salt (Sigma-Aldrich, China), Avicel® PH-200 (FMC Biopolymer, United Kingdom), Carrageenan, κ + λ (Sigma-Aldrich, USA).

1. Beauvra MA, Santelli RE, Oliveira EP, Villar LS and Escaleira LA. Response surface methodology (RSM) as a tool for optimization in analytical chemistry. Talanta. 2008; 76(5):965-977.
2. https://www.ioe.tuwien.ac.at/phd/planen/node82.html
3. https://www.itl.nist.gov/div898/handbook/pri/section3/pri3361.htm

Methodology

Twenty-four excipients were selected in the initial stage.

Then based on the preliminary studies only five were chosen that had the most favourable properties which led to a reduction of particle size without loss of flowability while maintaining a low *span* and X_{50}.

The optimal milling conditions were established using circumscribed central composite design evaluating the effect of milling time and speed on the responses (particle size, particle size distribution).

Factors for each excipient		Outcomes
Size of balls	2 mm, 5 mm, 10 mm	X_{90}, X_{50}, X_{10}, *span*
Speed	100 - 400 rpm'	
Time	15 to 45 min (rounded)	

To calculate the prediction or optimal conditions of milling process, a quadratic model with two factors was used:

$$Y = \beta_0 + \beta_{11} * X_1^2 + \beta_{22} * X_2^2 + \beta_{12} * X_1 * X_2 + \beta_1 * X_1 + \beta_2 * X_2$$

where y is response, X_1 and X_2 are the factors speed of milling and time of milling and β denote regression coefficients characterizing an intercept (β_0), the main (β_1, β_2), quadratic (β_{11}, β_{22}) and the interaction (β_{12}) effects.

The prediction model and optimum values were calculated using Matlab 2021b software. As the responses four variables were evaluated (X_{10}, X_{50}, X_{90} and *span*).

The optimal conditions of milling for the selected material were defined for each material and varied depending on its type and properties, which have been analyzed then by SEM and DSC.

Results

Despite the fact that four outcome parameters were evaluated (*span*, X_{10}, X_{50} and X_{90}), the most interesting we consider the variable X_{90} as well as the *span*. Since the differences in the span turned out to be not significant (only small changed in the nature of distribution were observed after milling), only X_{90} was further considered as a representative parameter.

For three out of five excipients the best size of balls was indicated as 5 mm balls. For the rest two 2 mm and 10 mm balls gave the better outcome.

Material	Size of the balls		
	2 mm	5 mm	10 mm
Calcium alginate	400 rpm' 45 minutes 53.34	**400 rpm'** **45 minutes** **22.81**	400 rpm' 45 minutes 54.64
Alginic acid	**400 rpm'** **45 minutes** **10.37**	392 rpm' 45 minutes 17.13	400 rpm' 45 minutes 24.12
MCC 200	312 rpm'* 36 minutes 58.70	329 rpm' 45 minutes 46.08	**354 rpm'** **45 minutes** **33.36**
Carrageenan	400 rpm' 45 minutes 107.33	**400 rpm'** **45 minutes** **55.15**	400 rpm' 45 minutes 98.85
HPMC	400 rpm' 27 minutes 230.54	**400 rpm'** **45 minutes** **128.79**	400 rpm' 45 minutes 166.23

Conclusion

- Using the QbD approach, we demonstrated how to optimize the milling process of pharmaceutical powders and assessed the effect of the key factors on the material;
- CCDs are very efficient designs, providing much information on experiment variable effects in a minimum number of runs with high quality predictions over the entire design space.

The study was supported by the by the Funding Agency of Charles University under Grant No. 268120/2020, by SVV 260547, and by Dean's Fund.

6.4 Grant projects

Principal investigator

- Rector's Mobility Fund; 2021; Grant number: FM/a/2020-2-035
- Research program Development and Study of Drugs (PROGRES Q42), category A

Team member

Grant Agency of Charles University; 2019-2022; Grant number: 70119/2019; Title of project: *Development of colon-targeted liquid-solid systems for the local therapy of inflammatory bowel diseases*

Grant Agency of Charles University; 2019-2022; Grant number: 268120/2023; Title of project: *Increase in drug-carrier surface interactions as a tool for improving the dissolution rate of poorly soluble drugs*

6.5 Scientific experience abroad

4-month laboratory training at Manufacturing and Materials research group, Department of Farmacy, University of Copenhagen (Prof. Jukka Rantanen), Denmark

7 LIST OF REFERENCES

Abd El-Monsef, M. M. E., & Seyam, M. M. (2011). CDT-optimum designs for model discrimination, parameter estimation and estimation of a parametric function. *Journal of Statistical Planning and Inference*, *141*(2), 639–643. https://doi.org/10.1016/j.jspi.2010.07.010

Adesina, S. K., Wight, S. A., & Akala, E. O. (2014). Optimization of the fabrication of novel stealth PLA-based nanoparticles by dispersion polymerization using D-optimal mixture design. *Drug Development and Industrial Pharmacy*, *40*(11), 1547–1556. https://doi.org/10.3109/03639045.2013.838578

Ahmed, O. A. A., Kurakula, M., Banjar, Z. M., Afouna, M. I., & Zidan, A. S. (2015). Quality by design coupled with near infrared in formulation of transdermal glimepiride liposomal films. *Journal of Pharmaceutical Sciences*, *104*(6), 2062–2075. https://doi.org/10.1002/jps.24448

Ait-Amir, B., Pougnet, P., & El Hami, A. (2015). Meta-model development. In *Embedded Mechatronic Systems* (Vol. 2, pp. 151–179). Elsevier Inc. https://doi.org/10.1016/b978-1-78548-014-0.50006-2

Antony, J. (2014). Understanding Key Interactions in Processes. In *Design of Experiments for Engineers and Scientists* (pp. 19–32). Elsevier. https://doi.org/10.1016/b978-0-08-099417-8.00003-1

Antony, J., Coleman, S., Montgomery, D. C., Anderson, M. J., & Silvestrini, R. T. (2011). Design of Experiments for non-manufacturing processes: Benefits, challenges and some examples. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *225*(11), 2078–2087. https://doi.org/10.1177/0954405411395857

Badawi, M. A., & El-Khordagui, L. K. (2014). A quality by design approach to optimization of emulsions for electrospinning using factorial and D-optimal designs. *European Journal of Pharmaceutical Sciences*, *58*(1), 44–54. https://doi.org/10.1016/j.ejps.2014.03.004

Badawy, S. I. F., Narang, A. S., Lamarche, K. R., Subramanian, G. A., Varia, S. A., Lin, J., Stevens, T., & Shah, P. A. (2016). Integrated Application of Quality-by-Design Principles to Drug Product Development: A Case Study of Brivanib

Alaninate Film-Coated Tablets. *Journal of Pharmaceutical Sciences*, *105*(1), 168–181. https://doi.org/10.1016/j.xphs.2015.11.023

Beg, S., Swain, S., Rahman, M., Hasnain, M. S., & Imam, S. S. (2019). Application of Design of Experiments (DoE) in Pharmaceutical Product and Process Optimization. In *Pharmaceutical Quality by Design: Principles and Applications* (pp. 43–64). Elsevier. https://doi.org/10.1016/B978-0-12-815799-2.00003-4

Beg Sarwar. (2021). Design of Experiments for Pharmaceutical Product Development. In *Design of Experiments for Pharmaceutical Product Development*. Springer Singapore. https://doi.org/10.1007/978-981-33-4717-5

Bhattacharya, S. (2021). Central Composite Design for Response Surface Methodology and Its Application in Pharmacy. *Response Surface Methodology in Engineering Science*. https://doi.org/10.5772/INTECHOPEN.95835

Boonpuang, R., Mongkolwongroj, M., Sakulkalavek, A., & Sakdanuphab, R. (2020). Empirical Modeling and Optimization of Laser Bending Process Parameters using the Central Composite Design Method for HDD Slider PSA/RSA Adjustment. *Lasers in Manufacturing and Materials Processing*, *7*(3), 290–304. https://doi.org/10.1007/s40516-020-00122-2

Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). Statistics for experimenters. *Wiley series in probability and statistics*. Hoboken, NJ: Wiley.

Box, J. F. (1922). *R. A. Fisher and the Design of Experiments, 1922-1926* (Vol. 34, Issue 1).

Cafaggi, S., Leardi, R., Parodi, B., Caviglioli, G., & Bignardi, G. (2002). *An example of application of a mixture design with constraints to a pharmaceutical formulation*. www.elsevier.com/locate/chemometrics

Campisi, B., Chicco, D., Vojnovic, D., & Phan-Tan-Luu, R. (1998). Experimental design for a pharmaceutical formulation: optimisation and robustness 1. In *Journal of Pharmaceutical and Biomedical Analysis* (Vol. 18).

Casler, M. D. (2015). Fundamentals of experimental design: Guidelines for designing successful experiments. *Agronomy Journal*, *107*(2), 692–705. https://doi.org/10.2134/agronj2013.0114

Charoo, N. A., Shamsher, A. A. A., Zidan, A. S., & Rahman, Z. (2012). Quality by design approach for formulation development: A case study of dispersible tablets.

International Journal of Pharmaceutics, *423*(2), 167–178. https://doi.org/10.1016/j.ijpharm.2011.12.024

Chavez, P. F., Lebrun, P., Sacré, P. Y., De Bleye, C., Netchacovitch, L., Cuypers, S., Mantanus, J., Motte, H., Schubert, M., Evrard, B., Hubert, P., & Ziemons, E. (2015). Optimization of a pharmaceutical tablet formulation based on a design space approach and using vibrational spectroscopy as PAT tool. *International Journal of Pharmaceutics*, *486*(1–2), 13–20. https://doi.org/10.1016/j.ijpharm.2015.03.025

Das, U., & Mandal, S. (2023). *Formulation by Design: An Overview*. www.intechopen.com

Dejaegher, B., & Vander Heyden, Y. (2011). Experimental designs and their recent advances in set-up, data interpretation, and analytical applications. In *Journal of Pharmaceutical and Biomedical Analysis* (Vol. 56, Issue 2, pp. 141–158). https://doi.org/10.1016/j.jpba.2011.04.023

Descamps, M., & Willart, J. F. (2016). Perspectives on the amorphisation/milling relationship in pharmaceutical materials. In *Advanced Drug Delivery Reviews* (Vol. 100, pp. 51–66). Elsevier B.V. https://doi.org/10.1016/j.addr.2016.01.011

Dhoot, A. S., Fernandes, G. J., Naha, A., Rathnanand, M., & Kumar, L. (2019). *Design of experiments in pharmaceutical development*.

Dong, X., Mattingly, C. A., Tseng, M., Cho, M., Adams, V. R., & Mumper, R. J. (2009). Development of new lipid-based paclitaxel nanoparticles using sequential simplex optimization. *European Journal of Pharmaceutics and Biopharmaceutics*, *72*(1), 9–17. https://doi.org/10.1016/j.ejpb.2008.11.012

El-Sayed, H. M., Abdellatef, H. E., Hendawy, H. A. M., El-Abassy, O. M., & Ibrahim, H. (2023). DoE-enhanced development and validation of eco-friendly RP-HPLC method for analysis of safinamide and its precursor impurity: QbD approach. *Microchemical Journal*, *190*. https://doi.org/10.1016/j.microc.2023.108730

Felix Oliver Lindner, P., & Hitzmann, B. (2006). Experimental design for optimal parameter estimation of an enzyme kinetic process based on the analysis of the Fisher information matrix. *Journal of Theoretical Biology*, *238*(1), 111–123. https://doi.org/10.1016/j.jtbi.2005.05.016

Ferreira, G. N., Silva, M. G. R., Fraga, A. G. M., da Silva, L. C. R. P., Lira, L. M., Rodrigues, C. R., Castro, H. C., de Sousa, V. P., & Cabral, L. M. (2014). Preparation and scale up of extended-release tablets of bromopride. *Brazilian Journal of Pharmaceutical Sciences*, *50*(2), 291–300. https://doi.org/10.1590/S1984-82502014000200008

Goh, H. P., Heng, P. W. S., & Liew, C. V. (2018). Comparative evaluation of powder flow parameters with reference to particle size and shape. *International Journal of Pharmaceutics*, *547*(1–2), 133–141. https://doi.org/10.1016/j.ijpharm.2018.05.059

Gregor, M., Grznár, P., Mozol, Š., & Mozolová, L. (2023). Design of simulation experiments using Central Composite Design. *Acta Simulatio*, *null*, null. https://doi.org/10.22306/asim.v9i2.99

Gujral, G., Kapoor, D., & Jaimini, M. (2018). An updated review on design of experiment (doe) in pharmaceuticals. *Journal of Drug Delivery and Therapeutics*, *8*(3). https://doi.org/10.22270/jddt.v8i3.1713

Hassan, H., Adam, S. K., Alias, E., Affandi, M. M. R. M. M., Shamsuddin, A., & Basir, R. (2021). Central Composite Design for Formulation and Optimization of Solid Lipid Nanoparticles to Enhance Oral Bioavailability of Acyclovir. *Molecules*, *26*, null. https://doi.org/10.3390/molecules26185432

Hinkelmann, K., & Kempthorne, O. (2007). *Design and analysis of experiments, volume 1: Introduction to experimental design* (Vol. 1). John Wiley & Sons.

Huang, C. T., Xu, R. T., Chen, P. H., Jong, W. R., & Chen, S. C. (2020). Investigation on the machine calibration effect on the optimization through design of experiments (DOE) in injection molding parts. *Polymer Testing*, *90*. https://doi.org/10.1016/j.polymertesting.2020.106703

Jankovic, A., Chaudhary, G., & Goia, F. (2021). Designing the design of experiments (DOE) – An investigation on the influence of different factorial designs on the characterization of complex systems. *Energy and Buildings*, *250*. https://doi.org/10.1016/j.enbuild.2021.111298

Javed, M. N., Kohli, K., & Amin, S. (2018). Risk Assessment Integrated QbD Approach for Development of Optimized Bicontinuous Mucoadhesive

Limicubes for Oral Delivery of Rosuvastatin. *AAPS PharmSciTech*, *19*(3), 1377–1391. https://doi.org/10.1208/s12249-018-0951-1

Jaydip, B., Dhaval, M., Soniwala, M. M., & Chavda, J. (2020). Formulation and optimization of liquisolid compact for enhancing dissolution properties of efavirenz by using DoE approach. *Saudi Pharmaceutical Journal*, *28*(6), 737–745. https://doi.org/10.1016/j.jsps.2020.04.016

Jones, B., Allen-Moyer, K., & Goos, P. (2021). A-optimal versus D-optimal design of screening experiments. *Journal of Quality Technology*, *53*(4), 369–382. https://doi.org/10.1080/00224065.2020.1757391

Ju, H., Zhang, Z. J., Lin, C. X., Liu, Z. J., & Jiang, H. L. (2019). Design optimization and experimental study of coaxial powder-feeding nozzle in the laser cladding process. *IOP Conference Series: Materials Science and Engineering*, *474*(1). https://doi.org/10.1088/1757-899X/474/1/012008

Jung, Y., & Lee, I. (2021). Optimal design of experiments for optimization-based model calibration using Fisher information matrix. *Reliability Engineering and System Safety*, *216*. https://doi.org/10.1016/j.ress.2021.107968

Karahoda, R., Robles, M., Marushka, J., Stranik, J., Abad, C., Horackova, H., Tebbens, J. D., Vaillancourt, C., Kacerovsky, M., & Staud, F. (2021). Prenatal inflammation as a link between placental expression signature of tryptophan metabolism and preterm birth. *Human Molecular Genetics*, *30*(22). https://doi.org/10.1093/hmg/ddab169

Kumar, L., Sreenivasa Reddy, M., Managuli, R. S., & Pai K., G. (2015). Full factorial design for optimization, development and validation of HPLC method to determine valsartan in nanoparticles. *Saudi Pharmaceutical Journal*, *23*(5), 549–555. https://doi.org/10.1016/j.jsps.2015.02.001

Kumar, P. M., & Ghosh, A. (2017). Development and evaluation of silver sulfadiazine loaded microsponge based gel for partial thickness (second degree) burn wounds. *European Journal of Pharmaceutical Sciences*, *96*, 243–254. https://doi.org/10.1016/j.ejps.2016.09.038

Kushner, J., Langdon, B. A., Hicks, I., Song, D., Li, F., Kathiria, L., Kane, A., Ranade, G., & Agarwal, K. (2014). A quality-by-design study for an immediate-release tablet platform: Examining the relative impact of active pharmaceutical

ingredient properties, processing methods, and excipient variability on drug product quality attributes. *Journal of Pharmaceutical Sciences*, *103*(2), 527–538. https://doi.org/10.1002/jps.23810

Li, Z., Lu, D., & Gao, X. (2021). Optimization of mixture proportions by statistical experimental design using response surface method - A review. In *Journal of Building Engineering* (36). https://doi.org/10.1016/j.jobe.2020.102101

Liu, H., Rivnay, B., Avery, K., Myung, J. H., Kozak, D., Landrau, N., Nivorozhkin, A., Ashraf, M., & Yoon, S. (2020). Optimization of the manufacturing process of a complex amphotericin B liposomal formulation using quality by design approach. *International Journal of Pharmaceutics*, 585. https://doi.org/10.1016/j.ijpharm.2020.119473

Mamidi, H. K., Palekar, S., Nukala, P. K., Mishra, S. M., Patki, M., Fu, Y., Supner, P., Chauhan, G., & Patel, K. (2021). Process optimization of twin-screw melt granulation of fenofibrate using design of experiment (DoE). *International Journal of Pharmaceutics*, *593*. https://doi.org/10.1016/j.ijpharm.2020.120101

Maretti, E., Rustichelli, C., Romagnoli, M., Balducci, A. G., Buttini, F., Sacchetti, F., Leo, E., & Iannuccelli, V. (2016). Solid Lipid Nanoparticle assemblies (SLNas) for an anti-TB inhalation treatment—A Design of Experiments approach to investigate the influence of pre-freezing conditions on the powder respirability. *International Journal of Pharmaceutics*, *511*(1), 669–679. https://doi.org/10.1016/j.ijpharm.2016.07.062

Marushka, J., Brokešová, J., Ugo Ogadah, C., Kazemi, A., Duintjer Tebbens, J., & Šklubalová, Z. (2022). Milling of pharmaceutical powder carrier excipients: Application of central composite design. *Advanced Powder Technology*, *33*(12). https://doi.org/10.1016/j.apt.2022.103881

Marushka, J., Hurychová, H., Šklubalová, Z., & Tebbens, J. D. (2022). Flow Equations for Free-Flowable Particle Fractions of Sorbitol for Direct Compression: An Exploratory Multiple Regression Analysis of Particle and Orifice Size Influence. *Pharmaceutics*, *14*(8). https://doi.org/10.3390/pharmaceutics14081653

Montgomery, D. C. (2017). *Design and analysis of experiments*. John wiley & sons.

Mura, P., Furlanetto, S., Cirri, M., Maestrelli, F., Marras, A. M., & Pinzauti, S. (2005). Optimization of glibenclamide tablet composition through the combined use of

differential scanning calorimetry and D-optimal mixture experimental design. *Journal of Pharmaceutical and Biomedical Analysis*, *37*(1), 65–71. https://doi.org/10.1016/j.jpba.2004.09.047

Myers, R. H., Montgomery, D. C., & Anderson-Cook, C. M. (2016). *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons.

Nikaeen, G., Yousefinejad, S., Rahmdel, S., Samari, F., & Mahdavinia, S. (2020). Central Composite Design for Optimizing the Biosynthesis of Silver Nanoparticles using Plantago major Extract and Investigating Antibacterial, Antifungal and Antioxidant Activity. *Scientific Reports*, *10*, null. https://doi.org/10.1038/s41598-020-66357-3

Oimoen, S. (2019). *Classical Designs: Full Factorial Designs*. www.afit.edu/STAT.

Panzade, P., Shendarkar, G., Shaikh, S., & Rathi, P. B. (2017). Pharmaceutical Cocrystal of Piroxicam: Design, formulation and evaluation. *Advanced Pharmaceutical Bulletin*, *7*(3), 399–408. https://doi.org/10.15171/apb.2017.048

Pires, F. Q., Angelo, T., Silva, J. K. R., Sá-Barreto, L. C. L., Lima, E. M., Gelfuso, G. M., Gratieri, T., & Cunha-Filho, M. S. S. (2017). Use of mixture design in drug-excipient compatibility determinations: Thymol nanoparticles case study. *Journal of Pharmaceutical and Biomedical Analysis*, *137*, 196–203. https://doi.org/10.1016/j.jpba.2017.01.037

Politis, S. N., Colombo, P., Colombo, G., & Rekkas, D. M. (2017). Design of experiments (DoE) in pharmaceutical development. In *Drug Development and Industrial Pharmacy* (Vol. 43, Issue 6, pp. 889–901). Taylor and Francis Ltd. https://doi.org/10.1080/03639045.2017.1291672

Ritu, Verma, R., Budhwar, V., & Kaushik, D. (2023). Investigation of solid dispersion approach for the improvement of pharmaceutical characteristics of telmisartan using a central composite design. *International Journal of Applied Pharmaceutics*, *15*(5), 245–254. https://doi.org/10.22159/ijap.2023v15i5.47968

Sharma, R., & Sarojwal, A. (2022). Influence of planetary ball mill parameters on powder flowability of AlSi10Mg with niobium carbide using central composite design (CCD). *Advances in Materials Science and Engineering*, *2022*

Shariare, M. H., Altamimi, M. A., Marzan, A. L., Tabassum, R., Jahan, B., Reza, H. M., Rahman, M., Ahsan, G. U., & Kazi, M. (2019). In vitro dissolution and bioavailability study of furosemide nanosuspension prepared using design of experiment (DoE). *Saudi Pharmaceutical Journal*, *27*(1), 96–105. https://doi.org/10.1016/j.jsps.2018.09.002

Sprent, P. (1973). Frank Yates and Experimental Design-Reflections Inspired by his Selected Papers. In *Source: Journal of the Royal Statistical Society. Series D (The Statistician)* (Vol. 22, Issue 2). https://about.jstor.org/terms

Sreenivasa Reddy, M., Kumar, L., Attari, Z., & Verma, R. (2017). Statistical optimization of extraction process for the quantification of valsartan in rabbit plasma by a HPLC method. *Indian Journal of Pharmaceutical Sciences*, *79*(1), 16–28. https://doi.org/10.4172/pharmaceutical-sciences.1000196

Srinivas, N. S. K., Verma, R., Kulyadi, G. P., & Kumar, L. (2017). A quality by design approach on polymeric nanocarrier delivery of gefitinib: Formulation, in vitro, and in vivo characterization. *International Journal of Nanomedicine*, *12*, 15–28. https://doi.org/10.2147/IJN.S122729

Tomuta, I., & Leucuta, S. E. (2004). Use of experimental design for identifying the most important formulation and technological variables in pelletization by powder layering. In *J. DRUG DEL. SCI. TECH* (Vol. 14, Issue 3).

Tošenovský, J. (2010). *Vysoká škola báňská-Technická univerzita Ostrava: Plánování experimentů učební text.*

Vasiljevic, D., Djuris, J., Jakimenko, S., & Ibric, S. (2017). Application of the fractional factorial design in multiple W/O/W emulsions. *Journal of Dispersion Science and Technology*, *38*(12), 1732–1737. https://doi.org/10.1080/01932691.2016.1278551

Vraníková, B., Svačinová, P., Marushka, J., Brokešová, J., Holas, O., Tebbens, J. D., & Šklubalová, Z. (2021). The importance of the coating material type and amount in the preparation of liquisolid systems based on magnesium aluminometasilicate carrier. *European Journal of Pharmaceutical Sciences*, *165*. https://doi.org/10.1016/j.ejps.2021.105952

Wu, C. J., & Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*. John Wiley & Sons.

Yue, R. X., Qin, H., & Chatterjee, K. (2011). Optimal U-type design for Bayesian nonparametric multiresponse prediction. *Journal of Statistical Planning and Inference*, *141*(7), 2472–2479. https://doi.org/10.1016/j.jspi.2011.02.010

Zatloukal, Z., & Šklubalová, Z. (2012). Effect of orifice geometry on particle discharge rate for a flat-bottomed, cylindrical hopper. *Particulate Science and Technology*, *30*(4), 316–328. https://doi.org/10.1080/02726351.2011.573839

Zidan, A., Kotamarthy, L., Ramachandran, R., Ashraf, M., & O'Connor, T. (2022). Optimization of screw design for continuous wet granulation: A case study of metoprolol succinate ER tablets. *International Journal of Pharmaceutics*, *623*. https://doi.org/10.1016/j.ijpharm.2022.121964