UNIVERSITÀ DEGLI STUDI DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI TECNICA E GESTIONE DEI SISTEMI INDUSTRIALI
DEPARTMENT OF MANAGEMENT AND ENGINEERING

Corso di Dottorato in Ingegneria Economico Gestionale
Ph.D. Course in Management Engineering and Real Estate Economics
XXXI CYCLE

DESIGN OF EXPERIMENT IN PRODUCTION PROCESS INNOVATION

Coordinator: Ch.mo Prof. Cipriano Forza
Supervisor: Ch.mo Prof. Luigi Salmaso

Ph.D. candidate: Fabrizio Ronchi

Thesis submitted in 2018

A Margherita e ai miei genitori

"In God we trust; all others bring data"
W.E.Deming

Sommario

Nel suo famoso libro *Progettazione ed Analisi degli Esperimenti* [54], Montgomery descrive il Design of Experiment (DOE) come un esteso approccio ad un esperimento che parte dalla definizione e dall'enunciato del problema, attraversa la progettazione della fase sperimentale e lo studio delle possibili soluzioni, per chiudersi con le conclusioni e le raccomandazioni. In particolare, il DOE è riconosciuto essere un potente strumento che si basa sulla statistica per progettare ed analizzare gli esperimenti. Le potenzialità del DOE sono ben conosciute ed apprezzate tra gli studiosi. In alcuni campi le sue potenzialità sono riconosciute ed apprezzate anche dai professionisti. Per questo motivo c'è un uso esteso del Design of Experiment nel miglioramento della qualità dei processi industriali. Ad esempio, il DOE è uno degli strumenti del Six Sigma per aumentare la qualità dell'output di un processo.

Secondo la definizione fornita da Bisgaard [8], l'innovazione è l'intero processo di sviluppo ed alla fine commercializzazione di nuovi prodotti e servizi, di *nuovi metodi di produzione* o approvvigionamento, di nuovi metodi di trasporto o servizi di consegna, di nuovi modelli di business, nuovi mercati, o nuove forme di organizzazione. Mentre l'uso del DOE è ben diffuso nella sperimentazione industriale per il miglioramento della qualità e della robustezza dei processi, il vantaggio dell'uso del DOE per l'innovazione è fonte di dibattito tra gli studiosi e tra i professionisti.

L'idea di studiare l'uso del DOE per l'innovazione dei processi di produzione ha origine da questo dibattito. La ricerca è stata condotta secondo diverse prospettive. La prima prospettiva riguarda l'efficacia del DOE nel supportare e potenziare la fase di innovazione di un processo produttivo. Essa è evidenziata grazie ad un caso studio nel quale è stata sviluppata una strategia per innovare il processo di termoformatura per la produzione di un packaging funzionale. Il DOE ha favorito la capacità di innovazione permettendo una riduzione degli errori sistematici e delle distorsioni, una completa esplorazione dello spazio fattoriale, ed una riduzione del numero dei test. L'approccio tradizionale per il controllo della produzione nei processi di termoformatura è stato messo alla prova e sfidato. Il DOE ha permesso di identificare e superare la discrepanza tra i fattori di controllo in laboratorio e quelli nella linea di produzione.

Una seconda prospettiva è di taglio manageriale ed è stata quella della gestione del processo di innovazione. L'impatto positivo che l'adozione del DOE ha avuto sulla gestione del processo di innovazione viene qui mostrato per mezzo di un caso studio. Il DOE ha dato prova di essere utile fornendo appropriati strumenti ed impattando su cinque dimensioni tipiche dell'ambito manageriale. Nello specifico: capacità decisionali, integrazione, comunicazione, tempi e costi, e gestione della conoscenza.

Per quanto concerne l'analisi dei dati, sono stati studiati alcuni metodi di analisi non parametrici. Attraverso uno studio di simulazione sono stati confrontati alcuni recenti test univariati non parametrici in un piano fattoriale a due vie. Lo studio ha mostrato come l'efficacia dei metodi di analisi vari a seconda del data set da analizzare e dell'obiettivo dell'analisi. Di conseguenza non è emerso un unico approccio da utilizzare nella fase di progettazione dell'esperimento, ma bensì vari aspetti devono essere tenuti in considerazione simultaneamente. Una accurata scelta del test più adatto favorisce l'impatto positivo che il DOE ha sull'innovazione dei processi di produzione. Inoltre è stato sviluppato un nuovo approccio multivariato non parametrico basato sulla NonParametric Combination (NPC) applicata ai test Synchronized Permutation (SP) per i piani sperimentali a due vie e due livelli. Questo approccio si è rivelato essere un buono strumento per la statistica inferenziale quando sono violate le assunzioni della MANOVA. Un grosso vantaggio fornito dall'adozione di questo tipo di test è che hanno ottime performances nel caso di campioni non numerosi. Questo fatto riflette le frequenti necessità dei professionisti in ambito industriale dove ci sono limitazioni o risorse scarse per la campagna sperimentale. Inoltre, c'è una proprietà molto importante della NPC dei test SP che può essere sfruttato per aumentare la loro potenza: la cosiddetta *finite sample consistency*. Infatti, può essere osservato un aumento della potenza sotto ipotesi alternativa H_1 quando il numero delle variabili risposta aumenta ed il numero dei campioni rimane costante. Questo fatto può fornire un beneficio strategico considerando che, in molti problemi concreti, potrebbe essere più facile raccogliere più informazioni da una singola unità statistica che aggiungerne una di nuova al piano sperimentale. Le proprietà di questo test multivariato lo rendono un utile strumento quando si intende utilizzare il DOE per l'innovazione dei processi produttivi e si verificano alcune condizioni specifiche.

Abstract

In his famous book *Design and Analysis of Experiments* [54], Montgomery describes Design of Experiment (DOE) as a broad approach to an experiment, starting from the recognition of and statement of the problem, going through the experimental design and to the possible solution, ending to conclusion and recommendations. Specifically, DOE is known to be a powerful instrument based on statistics to design and analyze experiments. Potentiality of DOE is well known and appreciated among scholars. In some fields its potentiality is recognized and appreciated also by practitioners. That's why there is an extensive use of Design of Experiment in improvement of industrial process quality. For instance, DOE is one of the instruments of Six Sigma to improve the quality of the output of a process.

According to the definition given by Bisgaard [8], innovation is the complete process of development and eventual commercialization of new products and services, *new methods of production* or provision, new methods of transportation or service delivery, new business models, new markets, or new forms of organization. While the use of DOE is well spread in industrial experimentation to improve quality and robustness of processes, the advantage of using DOE for innovation is debated among scholars and among practitioners.

The idea of investigating the use of DOE for production process innovation arose from this debate. Different perspectives have been investigated. The effectiveness of DOE to support and enhance the innovation of a production process is highlighted by means of a case study in which a strategy to innovate a thermoforming process for the production of a functional packaging has been developed. DOE enhanced innovation capability allowing reduction of systematic errors and distortions, full exploration of factorial space, and reduction of number of tests. Traditional approach to production control in thermoforming process was challenged. DOE allowed to identify and overcome the mismatch between control factors in laboratory and in production line. Another perspective was the management of the innovation process. The positive impact on innovation process management of adoption of DOE is shown by means of a case study. DOE proved to be helpful providing proper instruments, and impacting on five dimensions typical of managerial field. Namely: decision making, integration, communication, time and cost, and knowledge management.

Concerning the data analysis, some nonparametric methods of analysis have been investigated. A simulation study was used to compare some advanced univariate nonparamentric tests in a crossed factorial design. The study revealed that certain methods of analysis perform better than others depending on the data set and on the objective of the analysis. As a consequence, there does not emerge a unique approach in the design phase of the experiment, but various aspects have to be taken into account simultaneously. A thoughtful choice of the most suitable test enhances the positive impact that DOE has on the innovation of a production process. Furthermore, a novel multivariate nonparametric approach based on Non-Parametric Combination (NPC) applied to Synchronized Permutation (SP) tests for two-way crossed factorial design was developed. It revealed to be a good instrument for inferential statistics when assumptions of MANOVA are violated. A great advantage given by the adoption of these tests is that they well perform with small sample size. This reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. Furthermore, there is an important property of NPC of SP tests that can be exploited to increase their power: the finite sample consistency. Indeed, an increase in rejection rate can be observed under alternative hypothesis H_1 when the number of response variables increases with fixed number of observed units. This could lead to a strategical benefit considering that in many real problems it could be easier to collect more information on a single experimental unit than adding a new unit to the experimental design. Properties of this multivariate test make of it a useful instrument when using DOE to innovate a production process and some specific conditions are verified.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A production process is a series of mechanical or chemical steps used to create an object. It is a system that has been designed for a specific purpose and it is based on existing technology, scientific knowledge and experience of engineers in the specific field or application. Generally speaking, to innovate a system we have first to understand how it works. Then innovation, to be achieved, requires to exit the paradigm under which people is used to operate, requires to explore new solutions, and requires to understand how the new system will work under the new conditions.

Observing a production process while it is in operation it's useful in order to understand how it works. However, to fully understand what happens to a system and to its output when a change occurs in the way the system is organized or in some input factors, something more than simply observing is needed. Factors have to be changed according to a design, and changes have to be controlled and planned so that cause-and-effect relationships between modified input factors and output can be identified. In other words, experiments on the system have to be conducted. Design of Experiments (DOE) is a powerful instrument based on statistics to design and analyze experiments. Designed experiments can help to determine which input variables are responsible for a certain changement in the output response, and they can lead to a model relating response and input variables. Such experimentally determined models are called *empirical models* and can be used for production process improvement or other decision-making.

DOE plays an important role in product realization and commercialization activities. The objective of its use in many cases may be to develop a robust process, that is when external sources of variability have a small or negligible impact on the process and on its output. My research is about the use of DOE for the innovation of production processes. Specifically, it concerns those companies and organizations that invest in systematic innovation, no matter if radical or incremental innovation. In this doctoral thesis I present the results of the research starting from the gaps of knowledge that emerged in the systematic literature review, going through the impact of the use of DOE on the innovation capability and on innovation management, and ending to the benefits of the use of advanced techniques of analysis of factorial designs.

The remainder of this chapter is organized as follows. Section 1.1 provides a brief background of the research. Section 1.2 illustrates the research questions and the methodology adopted. Finally, in Section 1.3 a description of the thesis structure and of the content of each chapter is outlined.

1.1 Background

The Figure 1.1 adapted from Jensen [39] shows where statistical engineering is in an organization adopting statistical thinking. The organization is the pyramid in the picture. Statistical thinking, according to the definition from the American Society for Quality's (ASQ) Statistic Division (2011) "is a philosophy of learning and action based on the following fundamental principles: all work occurs in a system of interconnected processes; variation exists in all processes; and understanding and reducing variation are keys to process improvement". Statistical thinking is at a strategic level in an organization. Statistical engineering is at a tactical level, just over the operational level with methods and tools. Statistical engineering according to the definition of Hoerl and Snee [35] "is the study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results". If we add the statistics field (the ellipse in Figure 1.1), Statistical engineering is like a horizontal slice, connecting statistical theory with statistical practice. Statistical engineering is the natural environment of Design of experiment. Design of experiment is not simply a tool implemented in a software, not even a procedure. Design of experiment is a methodology that can be approached and implemented in various ways.

In his famous book *Design and Analysis of Experiments* [54], Montgomery describes Design of experiment as a broad approach to an experiment, starting from the recognition of and statement of the problem, going through the experimental design and to the possible solution, to conclusion and recommendations. Montgomery writes: "To use the statistical approach in designing and analyzing an experiment, it is necessary for everyone involved in the experiment to have a clear idea in advance of exactly what is to be studied, how the data are to be collected, and at least a qualitative understanding of

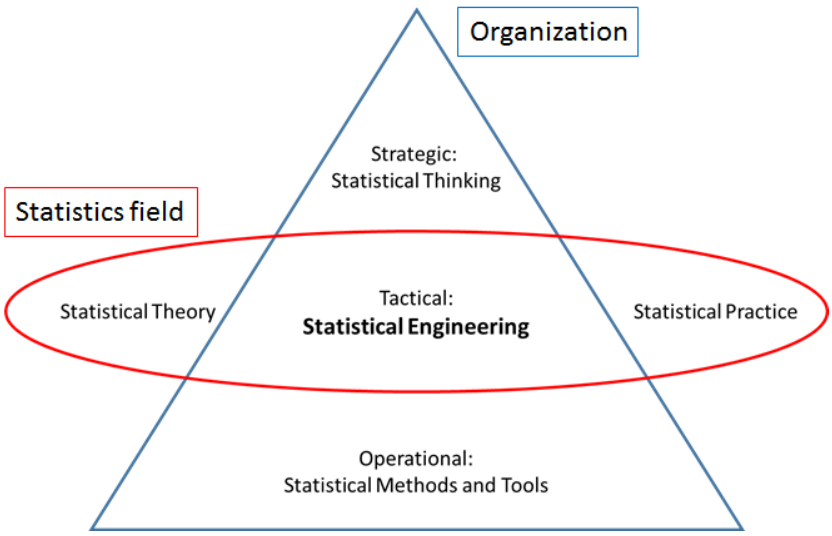


Figure 1.1: Statistical thinking and statistical engineering within the organization.

how these data are to be analyzed. An outline of the recommended procedure is [. . .]:"

1. Recognition of and statement of the problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of the data
7. Conclusion and recommendations

We have to note that the choice of the experimental design and the statistical analysis of data for which DOE is mostly known are just two intermediate steps even if DOE is often intended or addressed to as it was made up of only these two steps. The other steps are somehow underestimated in the common use and interpretation of DOE.

An experiment is a test or a series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reason for changes that may be observed in the output response [56]. As Box and Woodall [10] point out:"Many problems are complicated and contain many variables of interest. Experimentation is usually expensive. Statistical experimental design and, in particular, fractional factorial designs can minimize cost and maximize effectiveness". The application of experimental design techniques early in process development can lead to potential benefits as:

- Improved process yields
- Reduced variability and closer conformance to nominal or target requirements
- Reduced development time
- Reduced overall costs

Potentiality of DOE is well known and appreciated among scholars. In some fields its potentiality is recognized and appreciated by practitioners, that's why for instance there is an extensive use of Design of experiment in improvement of process quality. DOE is one of the instruments of Six Sigma to improve the quality of the output of a process. But the use of DOE for innovation is still debated. The idea to investigate the use of DOE for production process innovation arose from this debate.

Innovation is one of the leverages for competitiveness that companies are asked to use to stay in the market in the long term. Relevant scholars address innovation capability as one of the most important features a company should develop. Porter and Stern [67] said that the way companies can gain a business advantage today is "to create and commercialize new products and processes, shifting the technology frontier as fast as their rivals can catch up". W. E. Deming [17] some years earlier wrote that: "The moral is that it is necessary to innovate, to predict needs of the customer, and give him more. He that innovates and is lucky will take the market". And more recently Box and Woodall [10] said that: "Quality and efficiency cannot compete against the right innovation".

A definition of innovation is useful to better understand the purpose of the research. One clear definition is given by Bisgaard [8]: innovation is "the complete process of development and eventual commercialization of new products and services, *new methods of production* or provision, new method of transportation or service delivery, new business models, new markets, or new forms of organization". Jensen et al. [39] defined innovation as "the process of moving an initial invention or idea through research and development to the eventual market introduction". Focus of this study is production process.

Generally speaking, innovation involves changing the paradigm under which we operate to improve quality, efficiency, or the nature of a product or process. The issue of management of innovation processes is currently very actual, as innovations are widely recognized to be an important tool for increasing competitiveness of companies. Indeed, within the American Society for quality (ASQ), there has been an increasing focus on innovation with the recent creation of the ASQ Innovation Division. And Bisgaard [8] encouraged expanding or even rebranding quality engineering as innovation engineering.

One of many myths concerning innovation, is that innovation is the result of an idea that pops into one's head; that is, an epiphany or eureka moment. Rather, most innovations require a lot of work to refine and perfect the initial idea and to make it reality. This study is about companies or organizations that invest in systematic innovation, no matter if radical or incremental innovation.

1.2 Research Questions and Methodology

A systematic literature review updated at January 2017 has been conducted on the research topic *Design of Experiment in Production Process Innovation.* The literature review started from the initial idea about the research interest and allowed to map and assess the relevant intellectual territory leading to the discovery of some gaps of knowledge. The initial research interest thus evolved to the following justified research questions.

1. RQ1 What is the impact of adoption of Design of Experiment to innovate production processes? Does DOE support and enhance innovation of production processes?

2. RQ2 What is the impact of adoption of Design of Experiment on innovation process management?

3. RQ3 Do most recent data analysis techniques modify the impact of DOE on innovating production processes? Do they foster innovation?

Research questions 1 and 2 were faced by means of a case study. I was part of a team whose objective was to innovate the production system in order to achieve mass production capability of a product (a packaging for dishwasher detergent). My role was double. As expert in engineering and DOE, I was providing technical consultancy on the tools to adopt and how to use them. Namely, I was designing the fractional factorial experimental plans and analyzing data, furthermore I was involved in technical engineering issues. This quantitative research allowed to face research question 1 and it is widely described in Chapter 3. The second role was the role of *complete participant* in the adoption of DOE by the company as instrument to enhance innovation process. This research has been conducted according to the principles of qualitative research and is described in chapter 4. In a way consistent to the role of complete participant described by Macri and Tagliaventi [52], I was involved in the team and worked with the team. I could observe the behavior of people involved and study the impact of the adoption of DOE on innovation process management. I was thus able to understand the opinions of the members of the team and observe their evolution in time. Weekly meetings were planned in teleconference in order to coordinate the activities. Field notes and meeting reports were the most relevant instruments to trace the evolution of the innovation process management and the impact of DOE.

Research question 3 has been faced by means of two simulation studies. A simulation study consists in generating random data set according to certain models and distribution functions useful to evaluate the behavior of a method of analysis under specific conditions. In both the simulation studies data set were generated according to a linear additive cell mean model of a two-way crossed factorial design (factor A and factor B), i.e.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}. \tag{1.1}$$

where $i = 1,\ 2$ is the level of factor A; $j = 1,\ 2$ is the level of factor B; and $k = 1,\ \dots,\ n_{ij}$ is the k^{th} replicate for each factor level combination. In this setup, the general mean $\mu = 0$ and the interaction is given by the product of effects of the two factors. First simulation study was conducted to compare different univariate nonparametric methods of analysis and to assess their power (Chapter 5). Second simulation study was conducted to evaluate the performances of a novel nonparametric approach based on NonParametric Combination (NPC) applied to Synchronized Permutation (SP) tests for two-way crossed factorial designs (Chapter 6).

1.3 Stucture of the Thesis

This dissertation is structured in such a way that each chapter is independent and can be read by its own without necessarily reading other chapters. Starting from the literature review (Chapter 2), I move to the case study that was helpful to answer to the first (Chapter 3) and to the second (Chapter 4) research questions. Data set analyzed in the case study violated typical assumptions of parametric methods of analysis. Therefore, I had to resort to nonparametric tests both in the univariate and multivariate case. Most important, I had to chose appropriate tests and design the experiments accordingly. In Chapters 5 (univariate case) and 6 (multivariate case) I face the third research question. Two simulation studies give advice on the choice of non parametric method to be used depending on the data set and prove the effectiveness of novel method for multivariate analysis. Here below, a brief description of the content of each chapter is given.

Chapter 2 - Literature Review A systematic literature review has been conducted on the research topic *Design of Experiment in Production Process Innovation* according to the steps described by Tranfield et al. [76]. Systematic reviews differ from traditional narrative reviews by adopting a replicable, scientific and transparent process. The initial stage is the scoping study where definition, clarification and refinement of topics are carried out. Then, thanks to the identification of keywords and search term in the scoping study, documents in official literature are searched and selected according to specific criteria. The analysis and study of the documents collected allows to identify gaps of knowledge that lead to the definition of the research questions.

Chapter 3 - Case Study: an Experimental Strategy An experimental strategy on innovation of thermoforming for functional packaging has been developed as result of a research whose objective was to develop a new thermoforming process to produce a new packaging releasing detergent in dishwasher at a well-defined moment during main wash.

Innovation in thermoforming is a complex challenge. Thermoforming process is affected by various controlled and uncontrolled factors. Definition of factors, operating ranges and a deep understanding of their impact on the final product is fundamental. Furthermore, innovating a process requires to exit the paradigm under which people is used to operate, so a full understanding of existing process has to be achieved in order to be able to explore new solutions. There is no evidence in scientific literature of use of DOE for innovation of thermoforming production processes for active packaging.

The structured approach to the problem was based on the principle of Design of Experiment. Design of experiment was applied starting from the choice of raw material to the final test in dishwasher. Three control factors were initially chosen by a team of expert in thermoforming, that are temperature, time and pressure. These factors are commonly used to control material sealing in production line. The experimental approach revealed larger impact with current common technology of other factors such as preheating temperature and line pace. Final result was the development of a procedure that could allow correlation of control factors levels to packaging performance for each tested material, and selection of raw material and of factors' levels combination according to the desired packaging performance by means of an advanced DOE.

DOE enhanced innovation capability allowing reduction of systematic errors and distortions, full exploration of factorial space, and reduction of number of tests. Traditional approach to production control in thermoforming process was challenged. DOE allowed to identify and overcome the mismatch between control factors in laboratory and in production line.

Chapter 4 - DOE and Innovation Process Management The issue of management of innovation process is currently an important topic for many companies. Innovation is one of the ways, and in some fields the most important, to achieve and maintain a competitive advantage. It is a continuous process that has to be managed according to some dimensions. Furthermore, it is a cumulative process that has to be managed creating a favorable environment to let grow incremental innovation and to let breakthrough innovation be free to manifest. Innovation process is a process of recognizing customer needs and market opportunities for innovative products, generating and elaborating innovative ideas, working with knowledge regarding innovation and with information, realizing innovative activities and innovative production system to ensure a successful extension of the innovative product or service to the customers [49].

DOE is a methodology whose adoption could impact on the innovation process performances. Anyway, it's not DOE performing the innovation but people. DOE can provide a very effective and efficient aid that leads to innovation, but people are the ones that will produce results. Management of innovation process is necessary. Nevertheless, DOE could be helpful to improve the management of the innovation process providing proper instruments for managing the process.

The qualitative research performed during the adoption of DOE for the development of a new packaging production process allowed to highlight the impact of DOE on innovation process management specific to the case analyzed. In this chapter learnings are outlined and summarized along five dimensions that have a general value in the managerial field. Namely: decision making, integration, communication, time and cost, and knowledge.

Chapter 5 - Univariate Nonparametric Methods The two-way two-levels crossed factorial design is a common design used in the exploratory phase in industrial experiments. This design allows to investigate the impact on response variable of each factor and of interaction between factors, thus allowing to assess whether factors or interaction are significant or not according to the assumed model.

The F test in the usual linear model for analysis of variance (ANOVA) is a common instrument to compare the means of observations grouped according to factor level combinations, but sometimes data sets do not satisfy the assumptions of parametric tests. When assumptions like normal distributions of errors and homoscedasticity are violated, nonparametric tests are powerful instruments to support data-based decisions [36].

A variety of nonparametric methods have been developed during recent years. In this study I focus on Constrained Synchronized Permutations (CSP) [3], Unconstrained Synchronized Permutations (USP) [3], Wald Type Permutation (WTP) [61], Aligned Rank Transform (ART) [31] and ANOVA-Type Statistic (ATS) [11],which are designed to address the same hypotheses.

Practitioners have to convert the objective of the experiment in terms of type I and type II error rates. According to the objective, different levels of power of the test are required. Expected power of a test should be taken into account starting from the design phase of the experiment. Power of common parametric tests has been widely investigated and many software implement functions to calculate the power of a test according to the data set to be analyzed and other parameters. It is well known that factors that impact the power of a parametric test are: α level, variance, factor effect (expected difference between means) and number of replicates. Number of replicates could be a strong constraint when experiment is expensive.

In this study I compare the power of CSP, USP, ART, WTP, ATS and F tests in a two-way two-levels balanced factorial design. I fix the level α at 0.5 and I assess the power (P_w) along three dimensions of data set: i) factor effect, ii) standard deviation, and iii) number of replicates. The objective is to assess the impact of the three dimensions on the power of the tests and to give useful information on the choice of the test. I consider both the homoscedastic and the heteroscedastic cases. Furthermore, in a smaller scale simulation, I investigate the performances of the test in a two-way two-level unbalanced design varying the factor effect both in the homoscedastic and heteroscedastic case.

The simulation study allowed to assess the power of some selected nonparametric methods by analyzing the same data set. Data have been generated using a linear additive model for a two-way two-levels design with interaction given by the product of factor level effects. Such model is common for practitioners in industrial experimentation.

The study reveals that certain methods of analysis perform better than others depending on the dataset and on the objective of the analysis. As a consequence, there does not emerge a unique approach in the design phase of the experiment, but various aspects have to be taken into account simultaneously. The three dimensions (factor effect, standard deviation, number of replicates) along which the investigation has been conducted have an impact on the power of the tests, such as non-normal and heteroscedastic errors have. Furthermore, the study allowed to point out some interesting results.

Some interesting findings are related to (i) the different sensitivity of the tests to the variation of the dimensions,(ii) the conditions under which some tests can't be used, (iii) the tradeoff between power and type I error, and (iv) the bias of the power on one main factor analysis due to presence of effect of the other factor.

Chapter 6 - Multivariate Nonparametric Methods In many industrial applications (and applied research fields) it is common the need to compare multivariate population obtained in advanced factorial designs. There are manufacturing processes where treatments or control factors in production processes impact on several relevant variables simultaneously. In these cases an overall test is useful to determine for instance whether there is a significant difference on final product or not. Observed data are usually analized using the multivariate analysis of variance (MANOVA) methods. Unfortunately parametric methods rely on assumptions such as multivariate normality and covariance homogeneity, but these prerequisites may be not realistic for several real problems. How to overcome the violation of MANOVA assumptions has been investigated and nonparametric methods for multivariate inferential tests have been developed.

In my research I propose a novel nonparametric approach based on Non-Parametric Combination (NPC) [62] applied to Synchronized Permutation (SP) tests [3] for two-way crossed factorial design assuming a linear additive model. Indeed, the linear additive model interpretation well adapts to the industrial production environment because of the way control of production machineries is implemented. This approach overcomes the shortcomings of MANOVA with the only mild condition of the data set to be analyzed taking values on a multi-dimensional distribution belonging to a nonparametric family of non-degenerate probability distributions. It well works with even only two levels per factor and a small sample size. The case of small sample size reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design.

Furthermore it allows to formulate test hypotheses in more familiar terms for practitioners such as factor effect size.

A simulation design with fixed factor effects δ and fixed variance σ of data set distributions have been performed in order to evaluate the rejection rate of the NPC applied to SP under alternative Hypothesis H_1 in the range of interest of significance levels $0 \leq \alpha \leq 0.1$, and in order to compare it with the classical MANOVA test.

The application of NonParametric Combination to Synchronized Permutation to analyze a multivariate two-way factorial design reveals to be a good instrument for inferential statistics when assumptions of MANOVA are violated. Simulation results show that NPC applied to USP and CSP gives high values of power (rejection rate) under alternative hypothesis H_1 both with independent and dependent response variables, and both with low number and high number of response variables compared to MANOVA. A great advantage given by the adoption of these tests is that they well perform with small sample size. This reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. Furthermore, there is an important property of NPC of SP tests that can be exploited to increase their power: the finite sample consistency. Indeed, an increase in rejection rate can be observed under alternative hypothesis H_1 when the number of response variables increases with fixed number of observed units. This could lead to a strategical benefit considering that in many real problems it could be easier to collect more information on a single experimental unit than adding a new unit to the experimental design.

Chapter 7 - Discussion and Conclusions The research conducted as PhD student and results presented in this dissertation let me be on the same side of those scholars that in recent scientific literature stand in general in favour of the adoption of Design of Experiment for innovation, and in particular in favour of the adoption of DOE for production process innovation. Each step of the research was conducted with the aim of answering to the research questions that emerged from the systematic literature review. Nevertheless, a fil rouge ties together the case study by means of which I answered to the first and second research questions and the simulations studies by means of which I faced the third research question. The fil rouge consists in the challenges that I had to face in the analysis of the data set of the case study. This is typical when a research is conducted. You move on adapting step by step to the results and discoveries that you make. Main results and conclusions of the research are in this chapter summarized following the same structure of the dissertation.

Chapter 2

Literature Review

2.1 Introduction

A systematic literature review updated at January 2017 has been conducted on the research topic *Design of Experiment in Production Process Innovation* according to the steps described by Tranfield et al. [76]. The literature review started from the initial idea about the research interest and allowed to map and assess the relevant intellectual territory leading to the discovery of some gaps of knowledge. The initial research interest thus evolved to clear and justified research questions.

The research interest I started from was the impact of DOE on the R&D performances in innovating production processes. My personal experience as consultant suggested that there is a mismatch between the efforts to adopt and implement the use of DOE by a company and the needs of the company itself because of the misinterpretation of the support and benefits that use of DOE can allow. I wanted to investigate this aspect. Studying the impact of DOE on R&D performance could have given some hint on the reason of the mismatch. Furthermore, design of experiment is a very well known instrument by statisticians, so why isn't its use spread as wished for the innovation of production process? Generally speaking, the spread of the use of DoE for innovation is wished since nineties. Box in 1990 [9] wrote: "There are hundreds of thousands of engineers in this country, and even if the 2^3 factorial design was the only kind of experimental design they ever used, and even if the only method of analysis that was employed was to eyeball the data, this alone could have an enormous impact on the experimental efficiency, the rate of innovation and the competitive position of this country". Montgomery [55] suggests that, according to his experience, the reason could be that "most engineers have little exposure to design of experiments in their undergraduate academic training. It is typically part of another course that deal with several broad topics in engineering statistics, and DOE is only one item in the menu". So the poor knowledge that practitioners have of DOE could explain the lack of adoption by many companies.

Systematic reviews differ from traditional narrative reviews by adopting a replicable, scientific and transparent process. In the remainder of the chapter I present in details the steps of the literature review. The initial stage is the scoping study where definition, clarification and refinement of topics are carried out. Then, thanks to the identification of keywords and search term in the scoping study, documents in official literature are searched and selected according to specific criteria. The analysis and study of the documents collected allows to identify gaps of knowledge that lead to the definition of the research questions.

2.2 Background and scoping

First step was to deepen the topics related to the research interest in order to have a clear scope and to define boundaries of the review, and to understand the interest and the relevance of the study. Topics are *Design of experiment* and *Innovation*.

In his famous book *Design and Analysis of Experiments* [54], Montgomery describes Design of experiment as a broad approach to an experiment, starting from the recognition of and statement of the problem, going through the experimental design and getting to the possible solution, to conclusion and recommendations. Montgomery writes: "To use the statistical approach in designing and analyzing an experiment, it is necessary for everyone involved in the experiment to have a clear idea in advance of exactly what is to be studied, how the data are to be collected, and at least a qualitative understanding of how these data are to be analyzed. An outline of the recommended procedure is [...]:"

1. Recognition of and statement of the problem
2. Choice of factors, levels, and ranges
3. Selection of the response variable
4. Choice of experimental design
5. Performing the experiment
6. Statistical analysis of the data

7. Conclusion and recommendations

We have to note that the choice of the experimental design and the statistical analysis of data for which DOE is mostly known are just two intermediate steps even if DOE is often intended or addressed to as it was made up of only these two steps. The other steps are somehow underestimated in the common use and interpretation of DOE.

An experiment is a test or a series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify the reason for changes that may be observed in the output response [56]. As Box and Woodall [10] point out:"Many problems are complicated and contain many variables of interest. Experimentation is usually expensive. Statistical experimental design and, in particular, fractional factorial designs can minimize cost and maximize effectiveness". The application of experimental design techniques early in process development can lead to potential benefits as:

- Improved process yields
- Reduced variability and closer conformance to nominal or target requirements
- Reduced development time
- Reduced overall costs

Potentiality of DOE is well known and appreciated among scholars. In some fields its potentiality is recognized and appreciated by practitioners, that's why for instance there is an extensive use of Design of experiment in improvement of process quality. DOE is one of the instruments of Six Sigma to improve the quality of the output of a process. But the use of DOE for innovation is still debated.

Innovation is one of the leverages for competitiveness that companies are asked to use to stay in the market in the long term. Relevant scholars address innovation capability as one of the most important features a company should develop. Porter and Stern [67] said that the way companies can gain a business advantage today is "to create and commercialize new products and processes, shifting the technology frontier as fast as their rivals can catch up". W. E. Deming [17] some years earlier wrote that: "The moral is that it is necessary to innovate, to predict needs of the customer, and give him more. He that innovates and is lucky will take the market". And more recently Box and Woodall [10] said that: "Quality and efficiency cannot compete against the right innovation".

A definition of innovation is useful to better understand the purpose of the research and narrow down the literature review. One clear definition is given by Bisgaard [8]: innovation is "the complete process of development and eventual commercialization of new products and services, *new methods of production* or provision, new method of transportation or service delivery, new business models, new markets, or new forms of organization". Jensen et al. [39] defined innovation as "the process of moving an initial invention or idea through research and development to the eventual market introduction". Focus of this study is production process.

Generally speaking, innovation involves changing the paradigm under which we operate to improve quality, efficiency, or the nature of a product or process. The issue of management of innovation processes is currently very actual, as innovations are widely recognized to be an important tool for increasing competitiveness of companies. Indeed, within the American Society for quality (ASQ), there has been an increasing focus on innovation with the recent creation of the ASQ Innovation Division. And Bisgaard [8] encouraged expanding or even rebranding quality engineering as innovation engineering.

One of many myths concerning innovation, is that innovation is the result of an idea that pops into one's head; that is, an epiphany or eureka moment. Rather, most innovations require a lot of work to refine and perfect the initial idea and to make it reality. This study is about companies or organizations that invest in systematic innovation, no matter if radical or incremental innovation. This distinction anyway turned out to be relevant in some scholars' and practitioners' opinion about the use of Design of experiment for innovation.

During the preliminary research and study, two different positions emerged about use of DOE for Innovation. There are no doubts that statistics allow and facilitate decision making based on quantitative objective information, nevertheless there still is sometimes skepticism about the use of DOE for innovation mainly because DOE is seen as an instrument in the framework of quality control and improvement. Some scholars are in favor and some others are adverse. For instance, Doganaksoy stated that "statistically designed experimentation provides the foundation to guide large-scale innovation and development efforts through their multiple phases". While on the other hand Johnson said that "innovative environments thrive on useful mistakes and suffer when the demands of quality control overwhelm them [referring to the use of DOE and Six Sigma, Ed.]". We here provide a list of quotes and references that shows the two opposite positions on this theme.

Scholars and practitioners in favor of use of DOE for Innovation:

- Doganaksoy [39]: "Statistically designed experimentation provides the foundation to guide large-scale innovation and development efforts through their multiple phases".
- Anderson-Cook [39]: "Designed experiment can accelerate learning and evaluation of new ideas, which can streamline decision making about how to proceed, which leads to improved momentum for the innovative process".
- Box and Woodall [39]: "Design for Six Sigma methods can be used to develop both incremental and breakthrough innovations".
- Snee and Hoerl [39]: "A good place to start the list [of statistical tools for innovation, Ed.] are the tools of Lean Six Sigma and Design for Lean Six Sigma such as [. . .] design of experiment (DOE) [. . .]".
- O'Neill [39]: "Carefully designed experiments, with attention paid to sources of noise, potential interactions, measurement error, and feasibility (restrictions on randomization, maintaining manageable logistics), are the cornerstone of innovation".
- Montgomery [56]: "Design of experiment is viewed as part of a process for enabling both breakthrough innovation and incremental innovation, without which western society will fail to be competitive".

Scholars and practitioners adverse (or reporting adversity) to use of DOE for Innovation:

- Johnson [41]: "Innovative environments thrive on useful mistakes and suffer when the demands of quality control overwhelm them. [refering to the use of DoE and Six Sigma, Ed.]"
- Snee and Hoerl [39]: "An often stated concern about the use of statistical thinking and methods in any form, DOE, Lean Six Sigma, Lean DFSS, etc., is that it stifles creativity".
- Hargadon [27]: "Breakthrough innovations don't usually mix well with the pursuit of six sigma quality control, nor with those customers who just purchased your [statistical consulting companies, Ed.] last generation of products".
- Govinindarajan [33]: "The more you hardwire a company on total quality management, (the more) it is going to hurt breakthrough innovation".

Investigating the impact of the use of DoE on innovation make sense according to the fact that notable scholars and practitioners have opposite opinions on that. As consequence of the interest and opportunities that emerged in this preliminary work, a literature review has been conducted. Subject of literature review are studies about the effects and impact of adopting DOE's approach for innovation of production processes.

2.3 Database research and analysis of results

The database research was according to topics, keywords and logic that was defined during the preliminary research.

Topics are:

- Design of Experiment
- Innovation, with the restriction to the production

Keywords are:

- Design of experiment, Optimized experimental design and Fractional factorial design for Design of Experiment
- Process innovation and Process development for Innovation
- Production and Manufacturing for Production

In Figure 2.1 there is a sketch of the logic of the research in the database. The research was conducted on Scopus, Web of Science and Ebsco and was in Title, Keywords and Abstract. The result is given by the overlapping of the three areas (the circles in the figure) related to the mentioned topics. First circle on the top refers to DOE and there are numbers of documents found in Scopus, Web of Science and Ebsco. In the bottom left we have results from Innovation and in the bottom right we have production. Result is given by the overlapping in the center, that is 107 documents in Scopus, 21 in Web of Science and only 5 in Ebsco.

The selection criteria used on the database research are:

- Published in scientific peer-reviewed journals with H Index and SCImago Journal Rank
- No limitation of time
- Research in title, keywords and abstract

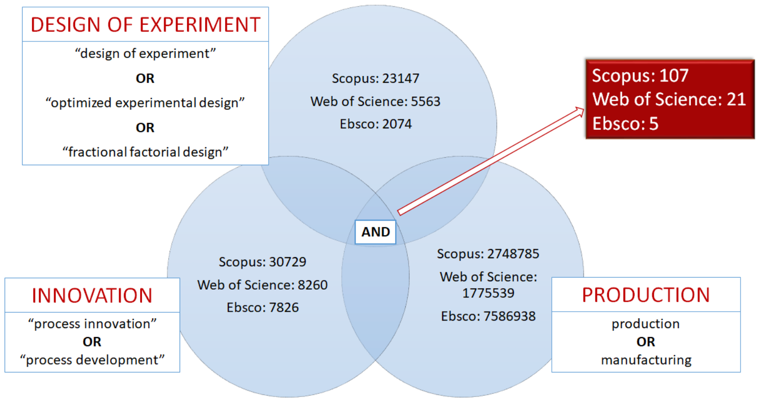


Figure 2.1: Results of research in electronic database: Scopus, Web of Science, Ebsco.

Table 2.1: Journals: H Index and SJR

Journal	H Index	SJR
International Journal of Operations and Production Management	94	2.198
Vaccine	142	2.044
Journal of Process Control	82	1.44
Organic Process Research and Development	78	1.411
Current Pharmaceutical Biotechnology	59	0.769
Biotechnology Progress	101	0.736
Engineering in Life Sciences	35	0.726
AAPS PharmSciTech	54	0.718
Total quality management	55	0.662
IEEE Transactions on Components, Packaging, and Manufacturing Technology	23	0.62
Advances in Biochemical Engineering and Biotechnology	69	0.527
Biotechnology and Applied Biochemistry	55	0.415
International Journal of System Assurance Engineering and Management	10	0.291
International Journal of Pharmaceutical Sciences Review and Research	16	0.193
BioPharm International	21	0.118

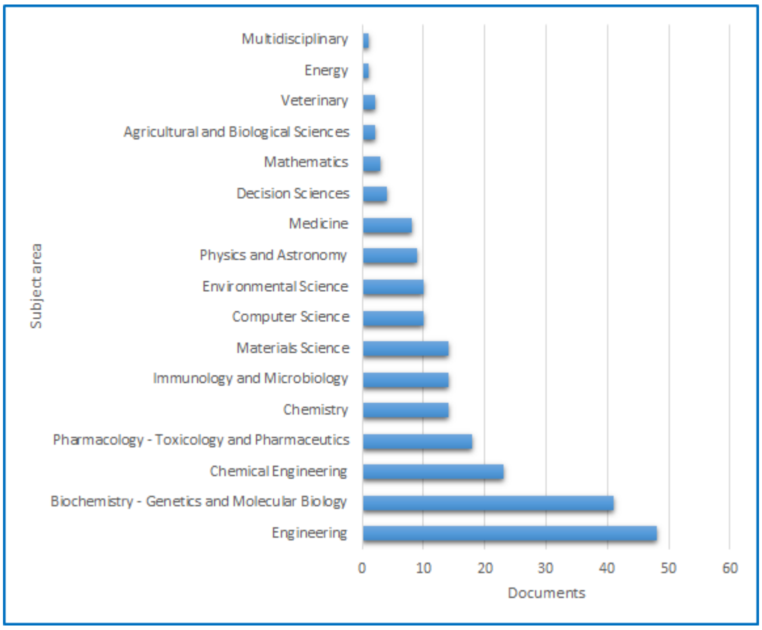


Figure 2.2: Documents per subject area.

In table 2.1 there is the list of journals with relative H Index and CSImago Journal Rank

An analysis along subject areas is provided in Figure 2.2. As we can see from the graph there are various subject areas, from Engineering and Biochemistry with the highest number of results, to Veterinary and Energy with the lowest, showing that, as we already said, DoE is a methodology whose features allows to apply it in very different fields of research. Number of documents per year (Figure 2.3) shows the recent increasing interest on the topic.

All the documents before being analyzed in detail had to pass a further selection according to some exclusion criteria. The exclusion criteria used for selection of documents are:

- No fits with research topic (from abstract reading) (e.g. focus on quality instead of innovation; e.g. machine learning)
- No English language (that means no international relevance)
- Published in conference proceedings

For the analysis of documents, it has been prepared a data extraction form to categorize papers and to collect data according to the dimension of

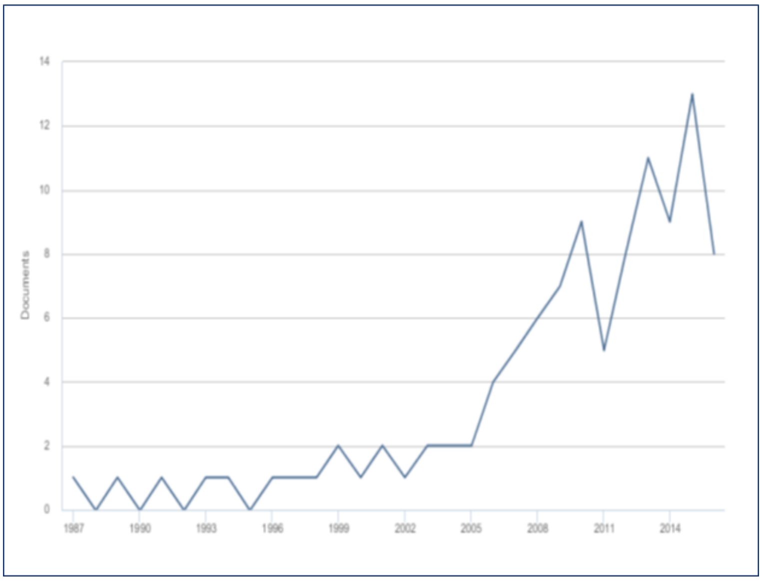


Figure 2.3: Documents per year, trend of publications.

analysis reported on Table 2.2. Some of the categories that have been used are generic like for instance General information (author, year of publication, etc.), Research type, Research methodology, etc., while some others are more specific and related to dimensions of the topics of my research, like for instance Type of DOE techniques, Type of data analysis, etc.

Table 2.2: Data extraction form

Dimension	Dimension's value or details
General information	Author/s, Year of publication, Title, Journal title
Research type	Exploratory, Theory building, Theory testing
Research methodology	Case study, Multiple case studies, Literature review, Mathematical model, Simulation, Survey, Action research, Conceptual
Type of analysis	Qualitative, Quantitative
Methods for collecting data	Experiment, Plant visit, Interview, Questionairre, Observation, Documents, etc.
Topic of research	Summary of the topic
Unit of analysis	Company, Function/s (name), Team, etc.
Methodology / approach in use ante DoE introduction	Type of methodology/approach (e.g. trial and error, etc.)
Type of DoE techniques	Full factorial, fractional factorial, optimization technique, etc.
Type of data analysis	Anova, Regression, Surface responce, Contour plot, etc.
DoE for quality improvement?	Yes, No
Implementation of DoE	E.g. Time, Resources, Sponsor, People involved, etc.
Impact of DoE on innovating performances	E.g. Time, Cost, Resources, Results reliability, etc.
Impact of DoE on organization	Describe (e.g. new roles/tasks in the company)
Short/Long term analysis?	Describe (e.g. Adoption of new approach in the company is long term, reach the goal is short term)
Generalizability of results	Yes, No (Why?)
Innovation	Yes, No

Analysis of data collected in the data extraction form highlighted some interesting aspects. The presence in papers of analysis of impact of DOE on innovation process according to different dimensions is reported in Figure 2.4. We can observe that the percentage of papers reporting the analysis of impact of DOE on time requested by the innovation process is slightly more than 50%, the percentage of the ones reporting the analysis of the impact on costs is 27% and on resources (from a general point of view) is 33%. Other dimensions are accounted for only 20% of the papers. Other dimensions are Robustness of process, Complexity reduction, and Effectiveness of methodology.

Then we have percentage of papers dealing with effects on organization as consequence of adoption of DOE, that is only 13% (Figure 2.5). And percentage of papers with short or long perspective of analysis of adoption of DOE (when present) where both perspectives are well represented (Figure 2.6).

Something more technical and very interesting is methodology in use ante DOE introduction. We have only 20% of documents mentioning it but not describing it (Figure 2.7).

DOE techniques found on the papers analyzed are listed below. These techniques are used along the various steps of DOE starting from recognition of and statement of the problem, going through choice factors and variables, and choice of experimental design, but not in the data analysis. Data analysis techniques are listed separately.

Type of DOE techniques:

- Cause and effect analysis
- Failure mode and effect analysis
- Full factorial design
- Taguchy design method
- Fractional factorial design
- Ishikawa diagram
- Face-centred central composite design
- Two level full factorial including center points
- ESS (environmental stress screening)
- Box-Behnken design

Figure 2.4: Presence in papers of analysis of impact of DoE on innovation process according to different dimensions.

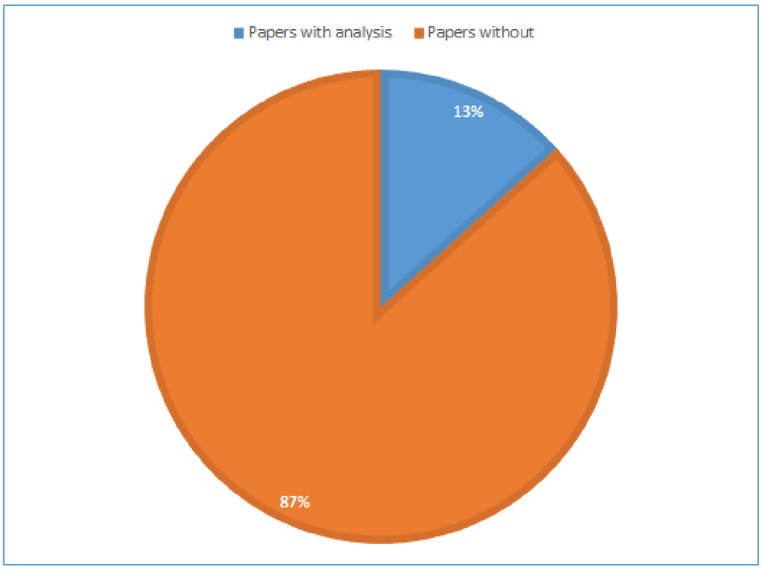


Figure 2.5: Percentage of papers dealing with effects on organization as consequence of adoption of DOE.

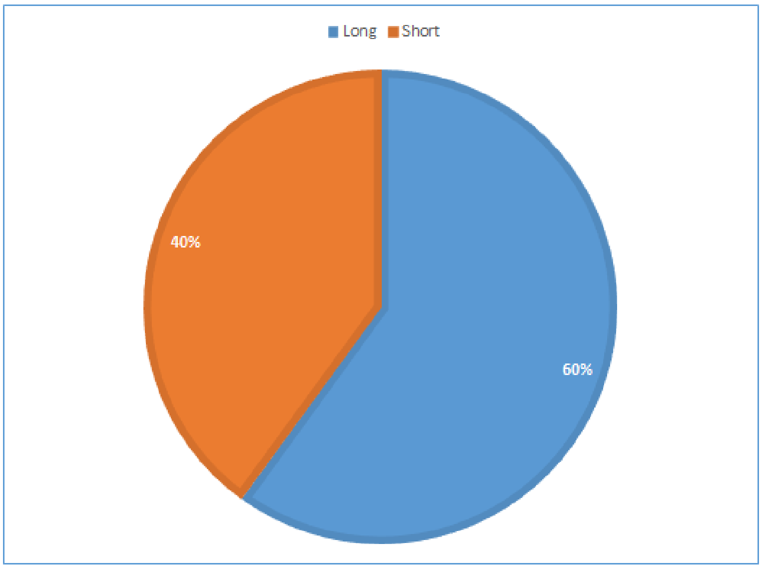


Figure 2.6: Percentage of papers with short or long perspective of analysis of adoption of DOE.

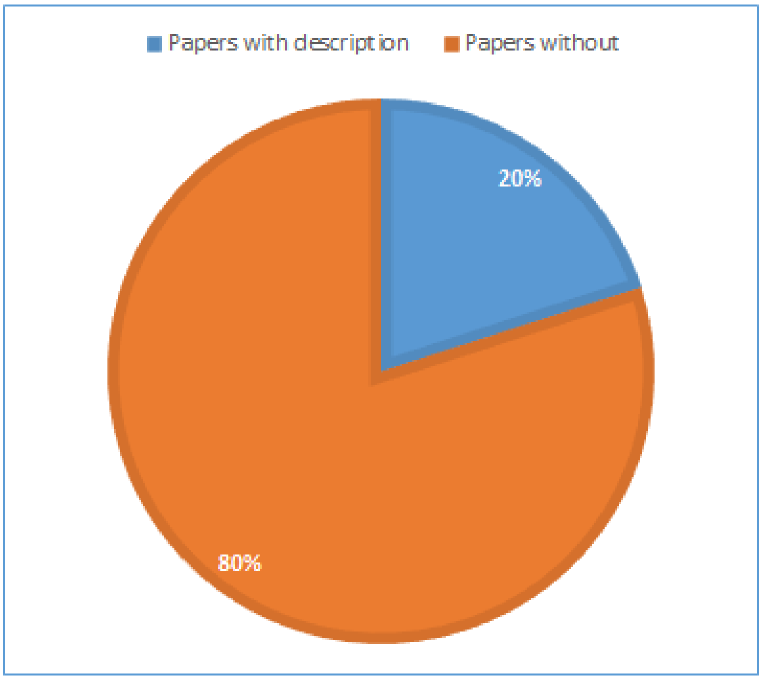


Figure 2.7: Percentage of papers reporting methodology in use ante DOE introduction.

- D-optimal design with center points

Type of data analysis techniques:

- t-test
- ANOVA
- Regression prediction models
- Gauge R&R
- Multivariate analysis
- ANOM (analysis of mean)
- Half normal plot
- Monte Carlo methods
- Response surface models
- Normal probability plot
- Second order polynomial fit
- Numerical optimization
- Hotelling's T^2
- Multiple linear regression
- Taguchy's quality loss function
- Contour plot

2.4 Gaps of knowledge and research questions

Evidence based analysis allowed to identify some gaps of knowledge:

1. GAP 1 In general there is lack of assessment of the impact of adoption of DOE on performances in innovating production processes. When present, it is mainly across three dimensions, that are time, costs and resources. A part from 'savings', evaluation across other dimension is poor and vague. There is lack of studies on if and potentially how use of DOE could support innovation capability and enhance innovation.

2. GAP2 There is lack of assessment of impact on firm organization. Adoption of DOE is often related to a single project. External consultancy is thus required. In these cases, there is not the adoption of DOE as a company tool.

3. GAP3 There is lack of evaluation of the efforts required to adopt or implement DOE. In general the adoption of DOE is not easy and the use is not spread as it could be.

4. GAP4 Many papers show how DOE has been used in a single case study in order to strengthen conclusions reliability, but studies do not show the benefit of DOE compared to other techniques. Comparative evaluations between techniques used before the adoption of DOE and DOE itself were not found.

5. GAP5 DOE techniques used for defining the problem, factors and levels and for designing of the experimental plan are high level and updated. Data analysis techniques are in general basic.

This lead to the following research questions:

1. RQ1 What is the impact of adoption of Design of Experiment to innovate production processes?

   RQ1.a Does DOE support and enhance innovation of production processes?

   RQ1.b What are the benefits of DOE compared to other techniques?

   RQ1.c What are the efforts required to adopt or implement DOE?

2. RQ2 What is the impact of adoption of Design of Experiment on innovation process management?

3. RQ3 Do most recent data analysis techniques modify the impact of DOE on innovating production processes? Do they foster innovation?

Chapter 3

Design of Experiment to Support the Innovation of Thermoforming Production Processes

3.1 Introduction

An experimental strategy on innovation of thermoforming for functional packaging have been developed as result of a research whose objective was to develop a new thermoforming process to produce a new packaging releasing detergent in dishwasher at a well-defined moment during main wash. Key part of the process and focus of this study is the sealing of polymeric film.

Innovation in thermoforming is a complex challenge. Thermoforming process is affected by various controlled and uncontrolled factors. Definition of factors, operating ranges and a deep understanding of their impact on the final product is fundamental. Furthermore, innovating a process requires to exit the paradigm under which people is used to operate, so a full understanding of existing process has to be achieved in order to be able to explore new solutions.

Previous studies on multilayer polymeric films investigated the influence of production parameters on their performances [13], the heat sealing properties of packages [82], and the influence of processing conditions on heat sealing behavior [37, 84]. Use of statistical approach in thermoforming is well spread. Design of Experiment (DoE) is used for process optimization [57, 47, 73], and for quality improvement [70], but there is no evidence in scientific literature of use of DoE for innovation of thermoforming production processes for active packaging.

The structured approach to the problem was based on the principle of

Design of Experiment. In fact, "The application of experimental design techniques early in process development can result in: improved process yields; reduced variability and closer conformance to nominal or target requirements; reduced development time; reduced overall costs" (Montgomery, D.C. 1997). Design of experiment was applied starting from the choice of raw material to the final test in dishwasher. Three control factors were initially chosen by a team of expert in thermoforming, that are temperature, time and pressure. These factors are commonly used to control material sealing in production line. The experimental approach revealed larger impact with current common technology of other factors such as pre-heating temperature and line pace. Final result was the development of a procedure that could allow selection of raw material and of factors' levels combination according to the desired packaging performance by means of an advanced DoE. There are several contributions and examples in scientific literature on techniques used in this study such as optimal design construction [59] and response surface [7].

This chapter illustrates the experimental strategy. The case study provides example of its development and help to illustrate steps and benefits of a structured experimental design. At the end the experimental strategy is formalized and a flowchart is provided.

Results presented in this chapter have been published [68].

3.2 The Product

The product under development is a single-dose bottle for detergent. It is made out of a polymeric film bent, formed and sealed. The bottle consists of two chambers. Polymeric material is sealed all around the bottle. Two different seals are required. First seal, from now on named Strong seal, is all around the bottle a part from the upper part. Second seal, from now on named Weak seal, is in the top of the two chambers. Strong seal is commonly easy to achieve and it is a common application of polymeric film for packaging. Strong seal has to guarantee resistance to shock during transportation, safe use for customers handling packaging, and permanent sealing in time. Weak seal is the challenge of this research. Weak seal has to open at a well-defined moment under external conditions that creates during washing cycle in dishwasher. It has to be weaker compared to strong seal that does not have to open under same conditions. Nevertheless, it has to resist to transportation shocks and guarantee safety of final customer. This means that a narrow window in weak seal performance has to be individuated and respected.

3.3 Case Study

3.3.1 Study of performances of materials at lab level

Market of polymeric film for thermoforming applications is huge and so is products offer from each supplier. Coupled PET/PE films in the market have different features, and selection is made according to the specific industrial application. In present specific case, novelty is given by the fact that same material is required to perform in two clearly different ways when sealed. How to measure final performance will be explained in next sections, but this preliminary observation is fundamental for the choice of material.

Objective of characterization is the description of features of material based on a certain response variable measured on seals obtained according to different configurations of production control factors. A team of expert of R&D department chose three control factors: i) Temperature of sealing jaws, ii) Time of sealing and iii) Force per cm^2 applied to the material. Sealing was performed controlling the three factors. Response variable is measured by a tensile strength test giving the force needed to open the seal tearing apart the two extremities. The idea is that the force needed to open the seal is representative of the final performance required to the packaging. The measure is obtained by a dynamometer.

Most important features of material according to the purpose of this study are:

1. *Differentiability between weak and strong seal.*
   Differentiability between weak and strong seal is fundamental. Response variable cannot vary within a small interval. It has to show enough difference between values observed at different factors levels combinations in order to achieve the control in production of weak and strong seal.

2. *Low variability of seal performance related to small fluctuation of production control factors.*
   Industrial production is in general affected by variability of input factors compared to laboratory experimental environment. To gain the control of production process, levels of control factors have to be chosen among those configurations that shows small effects on response variable as consequence of common fluctuations of input factors.

3. *Low variability of seal performance related to level of control factors.*
   Previous experiences of the team involved in the project revealed high variability of strength test at certain levels of control factors depending

on the material. In presence of such high variability, the phenomenon under investigation is out of control. Industrial production of packaging would be impossible at those factors levels combination that gives high and uncontrolled variability in response.

Test of Materials

Characterization of first candidate material (material A) were performed. In laboratory flat foils were sealed and tested. Sealing procedure was according to a protocol in order to reduce variability of response: flat foil were cut in squares 10 cm x 10 cm, randomly coupled and sealed always at the same distance from the border; sealing bars were steel made Teflon coated in order to avoid sticking to foil surface; temperature stabilization time of sealing bars was 10 minutes; once sealed, an overnight curing time have been respected in order to allow stabilization of polymeric bonds; 15 mm width stripes were cut out from the sealed squares and prepared to be tested; external stripes were excluded from analysis to avoid distortions given by border effects. There was a protocol for tensile strength test as well: threshold force before start = 0.1 N; displacement speed of grippers = 50 mm/min; initial gap between grippers = 35 mm. Both sealing and testing were performed by the same expert operator for the whole experimental design to avoid variance introduced by different operators. The result of tensile strength test is the maximum force measured to open the seal tearing apart the two extremities, from now on named Seal strength. Measure of seal strength value is in N/15 mm because the opening of the seal is orthogonal to the length that is 15 mm. In Figure 3.1 an example of a tensile strength test of sealed polymeric material.

Experimental design

At the beginning a one factor at time (OFAT) experiment allowed to identify operative ranges of factors for material A. Some ramps have been performed for each factor with the other two factors fixed at common levels for sealing, and ranges for factors were defined: Temperature 120 - 150 °C; Time 1 - 1,6 s; Force 20 - 60 N/cm^2. Operative ranges are specific for material A. Their definition should avoid inconsistent or inhomogeneous seal pattern that means too weak seal. At the opposite ranges definition should avoid even too strong seal that lead to material delamination that occurs when the two layers of PET and PE detach and material breaks but seal does not open.

Then an optimal factorial design with 3 factors and 4 levels each was

Figure 3.1: Example of a tensile strength test of sealed polymeric material. The test measures the force necessary to open the seal tearing apart the extremities of the stripe.

Table 3.1: Levels of control factors selected for the experimental design

Temperature		Time		Force per cm^2	
level code	value [°C]	level code	value [s]	level code	value [N/cm^2]
1	120	1	1.0	1	20
2	130	2	1.2	2	35
3	140	3	1.4	3	45
4	150	4	1.6	4	60

used to explore the response space in an effective and resource saving way. Levels of each factor have been selected in order to have almost homogenous intervals in the ranges. Table 3.1. Reduction of number of combination from the full factorial design (64 combinations) to the fractional factorial design (38 combinations) were achieved according to D-optimality criteria to minimize the variance in the regression coefficients of the fitted design model. The model selected includes terms up to order two, so that second order interactions can be estimated. Sequence of factors levels combination was random during specimens preparation. Number of replicates was 8. The number of replicates is precautionary in order to guarantee high power of inferential tests that could be necessary in this explorative phase and taking into account that this test does not need long time to be performed. When needed a sequential experiment is useful to zoom in the factors levels.

There are different options to design the experiment. Central Composite Designs are commonly used but best option for this study is Optimal Design to avoid problems with non-cuboidal regions [1]. In fact, preliminary experiments revealed that the factorial space defined according to selected ranges couldn't be explored entirely without falling into meaningless data regions, that is inconsistent seal pattern or material delamination. That's why an optimal factorial design [18, 40] with 3 factors and 4 levels each was used to explore the response space.

Data Analysis

The analysis of response variable aims to understand the impact of control factors on the most important features of the material, and to describe how response varies according to different factors levels combinations. Analysis of Variance of a full quadratic model allowed evaluating both main factors and interactions. A full quadratic model takes into account all linear terms, all squared terms, and all two-way interactions. A backward stepwise model selection procedure at a significance level $\alpha = 0.05$ has been applied to obtain the final model. Table 3.2.

Table 3.2: ANOVA table of the final model for material A.
T = Temperature, t = Time, F = Force per cm^2

Source of Variation		DF	Sum of Squares	Mean Square	F-Value	P-Value
Model		7	588.270	84.039	188.30	0.000
	T	1	351.874	351.874	788.44	0.000
	t	1	39.248	39.248	87.94	0.000
	F	1	4.310	4.310	9.66	0.002
	T^2	1	131.648	131.648	294.98	0.000
	t^2	1	4.054	4.054	9.08	0.003
	T*t	1	8.919	8.919	19.98	0.000
	T*F	1	5.655	5.655	12.67	0.000
Residual		296	132.102	0.446		
Total		303	720.372			

Table 3.3: Effects of the terms of the final model in coded units and Variance Inflation Factor.
T = Temperature, t = Time, F = Force per cm^2

Term	Effect	VIF
T	2.9388	1.04
t	0.9818	1.03
F	0.3412	1.04
T^2	-2.9797	1.01
t^2	-0.5319	1.05
T*t	-0.6427	1.04
T*F	-0.5334	1.07

According to P-Values the model and the three main factors are significant to explain the response variable. That is the means of the response variable are different at a confidence level of 95% between at least two factor levels. Only two interactions are significant. Interaction between Time and Force per cm^2 has not a significant impact on the seal strength. Coefficient of determination (R^2) of model is 81,66%. Analysis of residuals for ANOVA assumptions does not emphasize particular trends in data patterns, that is no violation of assumptions.

The impact of the terms of final model can be assessed. Table 3.3. To reduce the impact of non-orthogonal terms, the model was fitted in coded units. The effect of Temperature is much bigger than effect of Time, which in turn is bigger than effect of Force per cm^2. Low level of Variance Inflation Factor (VIF) shows absence of multicollinearity and confirms the goodness of the model.

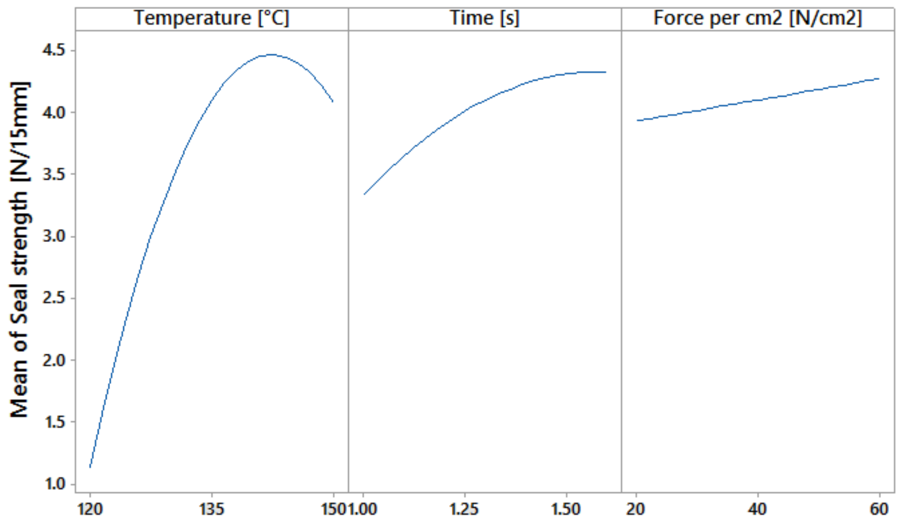


Figure 3.2: Main effects plot for Seal strength, fitted means: Mean of Seal strength VS Temperature, Time, and Force per cm^2

The model allows obtaining a nonlinear regression equation for Seal strength:

$$S = -154.60 + 2.014T + 18.96t + 1.286F - 0.006622T^2 - 2.955t^2 \\ - 0.0714T \cdot t - 0.00889T \cdot F$$

Where: S = Seal strength; T = Temperature; t = Time; F = Force per cm^2

The equation provides 'prediction' of response variable for those factors values that have not been tested during the experiment (only in the investigated ranges of factors values).

The effect on the response variable of each factor can be graphically analyzed, Figure 3.2. The analysis is based on the nonlinear regression equation. Temperature has a high impact on Seal strength. The range of observed values of Seal strength mean is wide (between 0.92 and 4.35 N/15 mm) at Temperature varying. There are steep slopes moving from a level of Temperature to the next one. There is a maximum at approx. 140 °C, then Seal strength decreases at 150 °C. Time effect plot has generally a positive slope with a flat zone at higher values. Force per cm^2 effect is the less evident. Range of observed values of Seal strength mean is between 2.93 and 3.41 N/15 mm at Force per cm^2 varying. Graphical analysis of interactions revealed general interaction between Temperature and Force per cm^2, and between the Temperature and Time only at high levels.

Nonlinear regression equation allows graphical analysis thanks to surface and contour plots. Surface plots show response variable in a 3D space for each couple of factors, while contour plots show response variable in a 2D space for each couple of factors. The remaining factor is fixed at a predefined level.

Surface plot gives a clear vision of the factors levels impact on response variable. Surface plots in Figures 3.3, 3.4 and 3.5 are just three examples of the plots analyzed. They show Seal strength versus each couple of factors holding the third at intermediate level. So, for instance, Figure 3.3 show Seal strength versus Time and Temperature, while Force per cm^2 is fixed at 40 N/cm^2 that by the way is not one of the levels of experimental design. This type of graphs provides interesting information about the impact of factors. In Figure 3.3 Flat area and Steep slope area are highlighted. Flat area individuates configurations of the two plotted factors levels that have a stable impact on response given the third level fixed. In fact, small variations of the factors don't show effects on response. In other terms the gradient of Seal strength in the flat area has a very low magnitude. This information is very important in industrial production to have full control of production process. Small fluctuations of input factors are common and they could have an impact on the stability of output. In case of input factors that show high fluctuations, configuration of factors levels should be chosen in a range giving a response surface as flat as possible. This information must be crossed with impact of the third factor taking into account at the same time the plots versus the three couple of factors. On the other hand, steep slope area reveals high impact of variation of factors. As rule of thumb, configuration of factors levels in the steep slope area should be chosen only when there is a good control of input factors in industrial production.

The three plots have been analyzed varying the third factor along the whole range. Flat area and steep slope area have been individuated and assessed for material A.

Contour plot is similar to surface plot but it is in two dimensions. In a contour plot two factors are in the axes, while the response variable is shown by different colored areas. The remaining factor is fixed at a predefined level. Contour plots in Figures 3.6, 3.7 and 3.8 are just three examples of the plots analyzed. They show Seal strength versus each couple of factors holding the third at intermediate level. In Figure 3.6 there is an example of how can be achieved differentiation between strong and weak seal by changing the level of Time and Temperature. Seal strength goes from lowest values in point A (approx. 0.5 N/15mm) to the highest in point B (approx. 4.2 N/15mm) holding Force per cm^2 at 40 N/cm^2. Areas contours are not straight but elliptical because of the combined effect of Temperature and Time. In Fig-

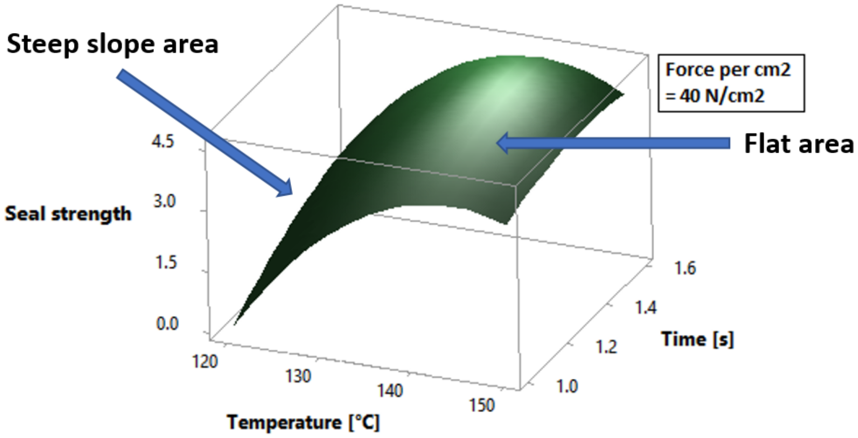


Figure 3.3: Surface plot of Seal strength VS Time and Temperature

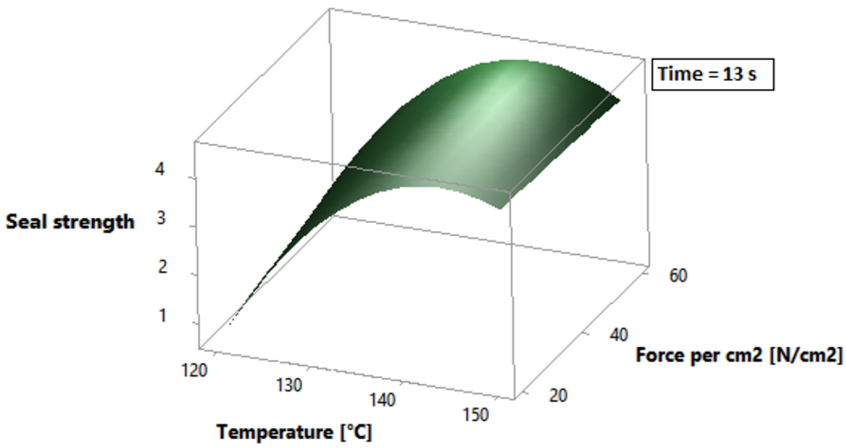


Figure 3.4: Surface plot of Seal strength VS Force per cm^2 and Temperature

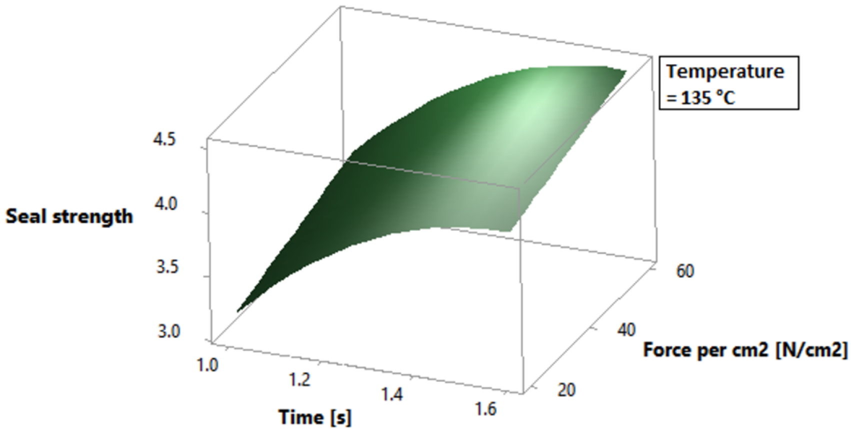


Figure 3.5: Surface plot of Seal strength VS Force per cm^2 and Time

ure 3.7 there is a different representation of the concept of flat area and steep slope area. In fact from 120 °C an increase of 5 °C temperature produces an increase in the Seal strength of more than 1 N/15 mm (steep slope area). At 140 °C same increase does not produce an increase in the Seal strength (flat area). Areas contours are almost straight, the higher impact of Temperature compared to Force per cm^2 is evident.

The three plots have been analyzed varying the third factor along the whole range. Differentiability between weak and strong seal has been individuated and assessed for material A.

Last graphical analysis according to the most relevant features of material is the one concerning data variability at certain factors levels combinations. Graph in Figure 3.9 show Seal strength distributions according to the 38 factors levels combinations of the experimental design. Response of material A has large variability at certain factors levels combinations. Variability could affect properties of final product given that a narrow window is expected for weak seal acceptable performance.

Comments on material characterization

Material characterization is performed to select candidate materials from market according to the most important features required for production. It allows studying differentiability between weak and strong seal, variability of seal performance related to small fluctuation of production control factors, and variability of seal performance related to level of control factors. As result of material characterization, a mathematical model to predict response of tensile strength test according to factors levels is developed. Material A

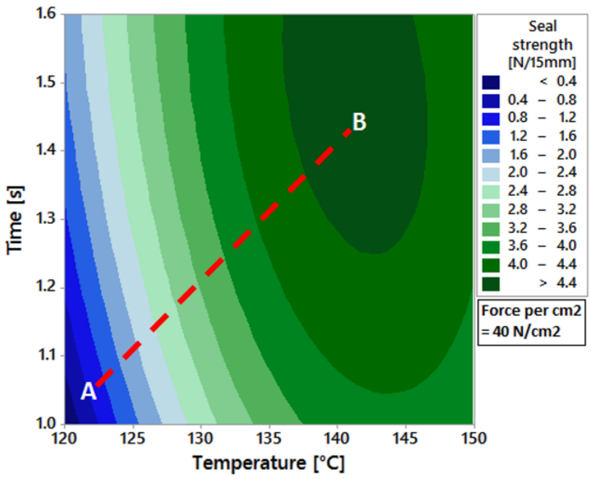


Figure 3.6: Contour plot of Seal strength VS Time and Temperature

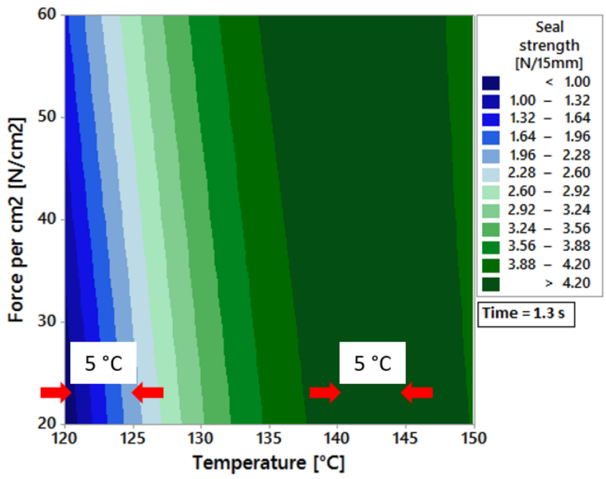


Figure 3.7: Contour plot of Seal strength VS Force per cm^2 and Temperature

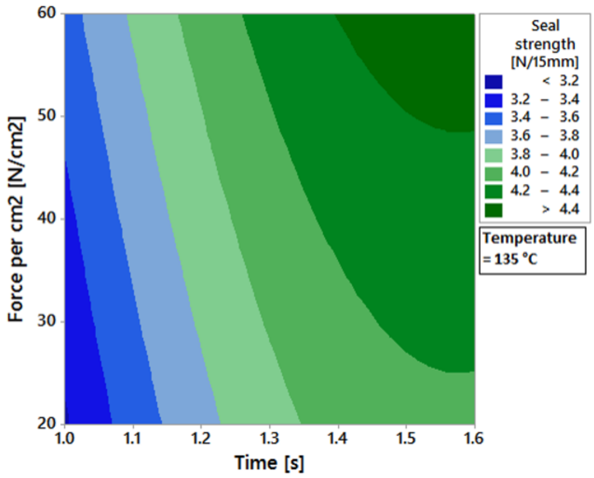


Figure 3.8: Contour plot of Seal strength VS Force per cm^2 and Time

provided a good example of the steps to be followed from identification of operative ranges to the analysis of experimental data. Material A revealed to be a good candidate for scale up to production line.

3.3.2 Study at production plant level

Production lines of packaging based on thermoforming of a polymeric film have a modular structure. Each module or station is devoted to a specific task, and assembly of stations is designed according to the desired packaging following general and established rules of thermoforming production. Film is in big rolls and production is continuous. Therefore, for instance, a production line devoted to production of a generic packaging for detergents could have this sequence of stations: loading station (feeding of line with film bent and coupled), pre-heating station, sealing station, forming station (air is blown into the packaging to give the final shape), filling station (packaging is filled with detergent), closing station (packaging is closed), labeling station (external label is applied to packaging), cutting station (final packaging is cut out from the line). Scheme of a production line devoted to production of a generic packaging for detergents is given in Figure 3.10.

The objective of the research is production of packaging from one material, and the challenge is to achieve weak and strong seals. Stations that could impact on seal performance are pre-heating station and sealing station. Pre-heating station prepares film before sealing. Film goes through the station, and two hot plates provide thermal energy. Temperature of pre-heating station is a control factor. Sealing station is composed by two faced molds

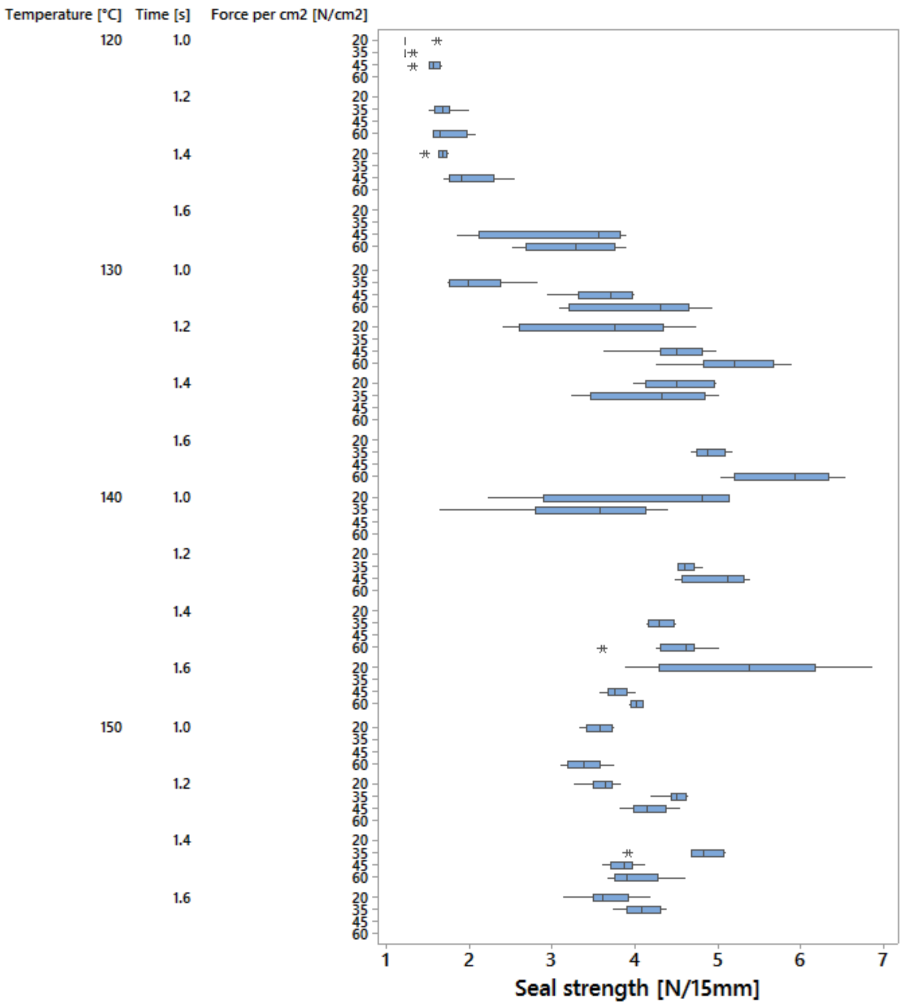


Figure 3.9: Boxplot of Seal strength according to factors levels combinations.

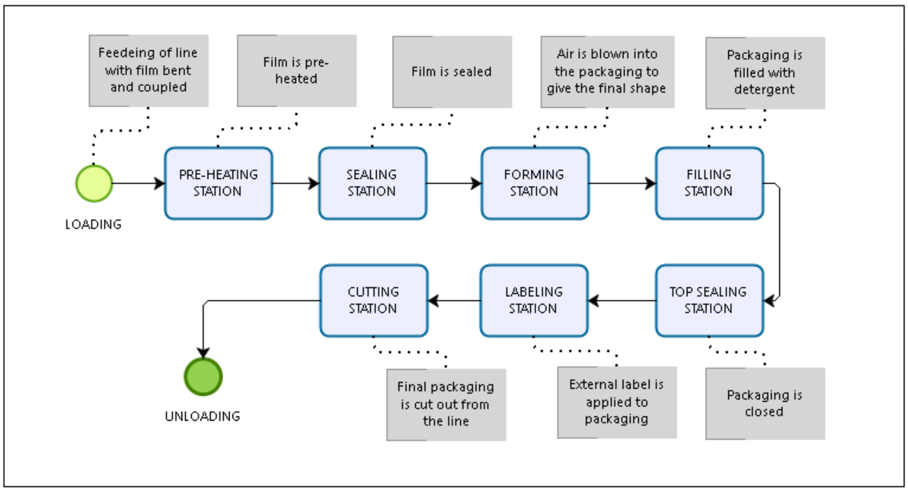


Figure 3.10: Scheme of production line

that close cyclically and seal the film. Temperature, time and force per cm^2 are control factors in sealing station.

Traditional approach to packaging production by thermoforming process consider temperature, time and force per cm^2 the most important factors to control seal features, while temperature in pre-heating station is used to control film performance in forming station. In fact, pre-heated material is soft and can be easily modelled by blowing air to achieve the final shape, then packaging is cooled down to fix the shape. Traditional approach make sense for traditional packaging, but production of weak seal for the new functional packaging challenged it.

Molds in sealing station are made of two different materials, one for the weak seal and one for the strong seal. Steel Teflon coated is used for strong seal jaws and polyether ether ketone (PEEK) is used for weak seal jaws. These two materials are embedded in one mold, and control of time and force per cm^2 is independent for the two materials so differentiation is allowed according to these two factors. Temperature control is unique for the whole mold and differentiation is achieved by the different thermal conductivity of the two materials (PEEK: 0.25 W/mK; Steel: approx. 50 W/mK). In static conditions a stable delta of temperatures between materials were observed and looked to be appropriate for seal differentiation. This technology in fact proved to be successful to produce packaging with different seals performance. A confirmatory experiment based on levels of factors used in laboratory for material A was thus performed. Control of weak seal jaws temperature was achieved thanks to direct measurement by thermocouples. Despite temperature, time and force per cm^2 conditions for weak sealing jaws were the same as the ones used for the experiments in the laboratory, results obtained in the pilot production line were completely different. Weak seal was inconsistent and material along strong seal started to burn because of high temperature settings. Weak sealing failed. The systematic experimental approach allowed to individuate the source of failure in thermal conductivity of PEEK. Further investigation showed how in a steady state during production, temperature of PEEK drops down compared to machine setting and that thermal energy transferred to material is insufficient for sealing. Laboratory experiment settings on flat film are not transferrable to pilot production line taking into account current production system.

As per results of lab experiments for characterization of material A, temperature has the biggest impact on seal performance compared to time and force per cm^2. New factors that could allow control of transferred thermal energy were investigated. Pre-heating station gives thermal energy to the film. Preliminary experiments allowed to select three main control factors for weak seal production in pilot line: i) Pre-heating station temperature; ii) Pace of the line; iii) Time of sealing. Pace of the line is measured in cycles per minute and it allows control of the time the film spends in the preheating station. The lower is the number of cycles per minute, the longer is the time spent by the film in the preheating station. Control of thermal energy transferred to the film is achieved by combination of pre-heating station temperature and pace of the line. Note that according to material characterization results, force per cm^2 was considered negligible as control factor in production line. Sealing was performed controlling the three factors.

The experimental response variable is measured by a burst test giving the pressure needed to open the seal by injecting air into the bottle. The idea is that the packaging shrinks when temperature in dishwasher raises because of the memory form effect of thermoformed polymeric material. The effect of shrinkage is the increase of internal pressure that at the end lead to the packaging opening. The pressure needed to open the seal is thus representative of the final performance required to the packaging.

The final objective is to produce a packaging releasing detergent in dishwasher at a well-defined moment during main wash. Response variable analysis should allow to select samples for dishwasher test in order to find a correlation between packaging performance in dishwasher and production factors levels, and in case of success to find the best factors setting for production. The reason of making a selection by pressure test before dishwasher test is that the latter is highly time consuming.

Test of product samples

Material A was used to produce batches of finite empty bottles according to an experimental design. In production line presence of uncontrolled factors and of fluctuations of controlled factors is higher compared to lab environment. Production was according to a protocol in order to reduce variability of response: production was in steady state; temperature stabilization time of pre-heating station was 10 minutes; sealing station temperature was fixed at 120 °C; force per cm^2 applied to weak seal was 410 N/cm^2. A curing time of at least 1 day have been respected in order to allow stabilization of polymeric bonds. There was a protocol for burst test as well: air is injected in a flat area of packaging far from seals; air is blown into the bottle according to a pressure ramp with steps of amplitude 0.02 bar; interval between steps is 5 seconds; pressure ramp starts at 0.05 bar and upper limit is 0.5 bar. Burst test were performed by the same expert operator for the whole experimental design to avoid variance introduced by different operators. The result of burst test is the minimum pressure needed to open the bottle. Pressure test device provides the relative pressure (over atmospheric pressure) as percentage of 1 bar.

The experiment is focused on weak seal. In fact, proven that material A allowed differentiability of performance in material characterization section, time and force per cm^2 for strong seal in sealing station were tentatively chosen in order to guarantee a minimum pressure at burst test of 0.5 bar.

Experimental design

At the beginning a one factor at time (OFAT) experiment allowed to identify operative ranges of factors for material A. Table 3.4. Some ramps have been performed for each factor with the other two factors fixed at common levels for production, and ranges for factors were defined: Pre-heating temperature 120 - 130 °C; Pace of the line 11 - 14 cycles per minute; Sealing time 1.6 - 2.2 s. Operative ranges are specific for material A. Defined ranges should avoid inconsistent or inhomogeneous seal pattern that means too weak seal. At the opposite ranges definition should avoid even too strong seal that prevent bottle opening in dishwater. Then an optimal factorial design with 3 factors and 4 levels each was used to explore the response space in an effective and resource saving way. Levels of each factor have been selected in order to have almost homogenous intervals in the ranges. Reduction of number of combination from the full factorial design (64 combinations) to the fractional factorial design (38 combinations) were achieved according to D-optimality criteria to minimize the variance in the regression coefficients of the fitted

Table 3.4: Levels of control factors selected for the experimental design

Pre-heating temperature		Pace of the line		Sealing time	
level code	value [°C]	level code	value [cycles/min]	level code	value [s]
1	120	1	11	1	1.6
2	123	2	12	2	1.8
3	127	3	13	3	2.0
4	130	4	14	4	2.2

design model. The model selected includes terms up to order two, so that second order interactions can be estimated.

Packaging is composed of two chambers. The chambers, from now on named left chamber and right chamber, were tested separately, that is for each packaging only left or right chamber were tested. In fact packaging is asymmetric and different geometry between the chambers could have an impact on weak seal performance. Number of replicates was 10, 5 to test left chamber and 5 to test right chamber. The number of replicates is precautionary in order to guarantee high power of inferential tests that could be necessary in this explorative phase.

Data Analysis

Data analysis was based on the same approach used for material characterization. General steps of analysis are here briefly summarized. For each chamber, Analysis of Variance of a full quadratic model allowed evaluating both main factors and interactions taking into account all linear terms, all squared terms, and all two-way interactions. A backward stepwise model selection procedure at a significance level $\alpha = 0.05$ has been applied to obtain the final models. The impact of the terms of final model was assessed. Variance Inflation Factor (VIF) was useful to exclude multicollinearity. The models allowed obtaining a nonlinear regression equation for response variable. Graphical analysis of the effect on the response variable of each factor and second order interactions was performed as well as analysis of surface and contour plots. Last but not least, graphical analysis of data variability at different factors levels combinations.

In general, an experiment performed in industrial plant presents a limit in data analysis compared to one performed in laboratory. Surface and contour plots graphical analysis in material characterization allowed to assess impact of fluctuation of factors. In fact, laboratory environment enable high control of factors, and variability observed at a certain factors levels combination is mainly related to uncontrolled factor, as for instance homogeneity of mate-

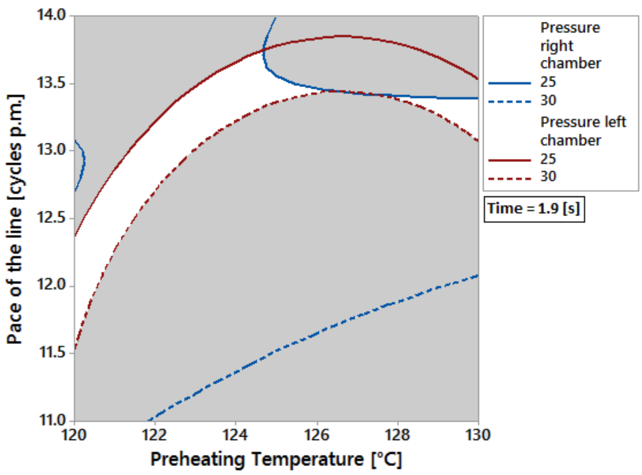


Figure 3.11: Overlaid contour plot of Pressure of right chamber and left chamber VS Pace of the line and Preheating temperature. Interval of response from 25% to 30%.

rial. In production line, control of factors is affected by typical variability of industrial environment and sensors are in general less sensitive. The effect of fluctuation of control factors is thus confounded with the effect of variability of uncontrolled factors. This limit should be taken into account when drawing conclusion from data analysis.

Range of observed values was from 0% to 50%, meaning that choice of factors levels allowed investigation of the whole interval of interest of response values. The three factors pre-heating station temperature, pace of the line and time of sealing are significant in the model of both chambers. Pace of the line is the most important factor when considering the impact on pressure value required to open the weak seal of both chambers of the samples. Pressure value increases at low levels of cycles per minute, while it decreases at high levels of cycles per minute.

One important result was that the test revealed a non-symmetric behavior of the left and right chamber of the bottle. Geometry influences the way the chambers open. Different distributions of response values were observed between the two chambers for many factors levels combinations. As consequence, nonlinear regression equations obtained from the models of left and right chambers were different. A well performing packaging should guarantee that detergent is delivered from the two chambers at around the same time during washing cycle. Therefore, those factors levels combinations that could allow similar performance were investigated.

Response contour plots of the two chambers were compared by overlap-

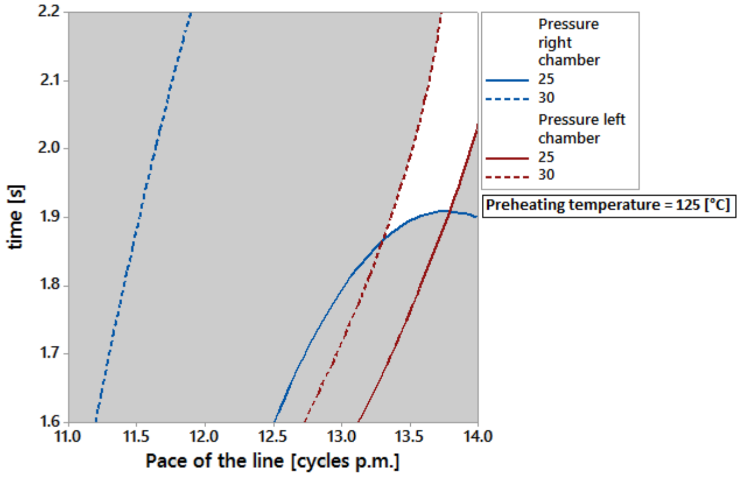


Figure 3.12: Overlaid contour plot of Pressure of right chamber and left chamber VS Time and Pace of the line. Interval of response from 25% to 30%.

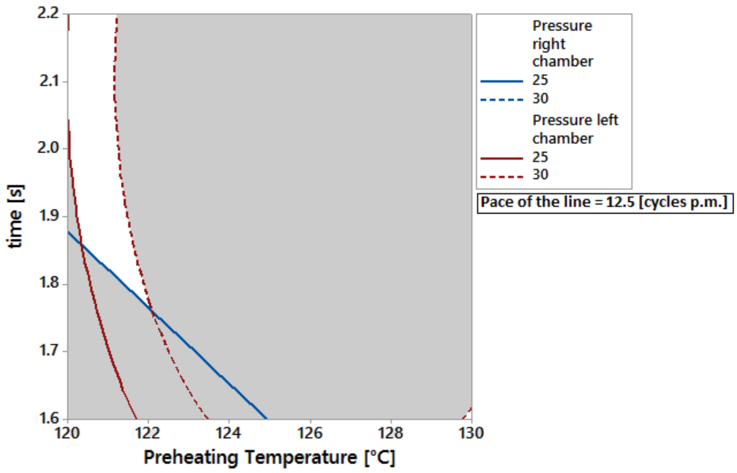


Figure 3.13: Overlaid contour plot of Pressure of right chamber and left chamber VS Time and Preheating temperature. Interval of response from 25% to 30%.

ping them two by two. The analysis of overlaid contour plots was performed for each couple of factors. Analysis allowed identification of areas where predicted responses were similar. Examples of overlaid contour plots in Figures 3.11, 3.12 and 3.13 are for interval of response from 25% to 30%. Intervals in the whole range of response have been investigated. Furthermore, the three plots have been analyzed varying the third factor along the whole range. The statistical model allowed to select those factors levels combinations that results in a similar behavior of the two chambers.

Comments on production in pilot line

Pilot line experiments are performed on those materials that show promising results according to material characterization. Material A provided a good example of the steps to be followed from identification of operative ranges for the three control factors in the production line to the analysis of experimental data. As result of experimental campaign, a statistical model to predict response of pressure test according to factors levels is developed. The objective is the selection of samples showing different performances at pressure test to perform the most significant test: the test in dishwasher. Selection must consider: i) coverage of different pressure test performances, ii) similar behavior of the two chambers, and iii) low variability of response according to factors levels combinations. Material A allowed selection of samples to be tested in dishwasher.

3.3.3 Study of final product prototypes

Performance of final product can be evaluated according two dimensions: i) time of opening during washing cycle, ii) complete vs partial release of detergents. These two dimensions are representative of the expectations of final customer. In fact, packaging has to deliver detergent on or before a defined threshold so that there is enough time for detergent for an effective cleaning. Furthermore, complete release of detergent is given as qualifier by marketing.

The two dimensions were investigated by a dishwasher test. Time of opening was established by means of sensors inside the dishwasher. The amount of delivered detergent was observed at the end of washing cycle. This test is time consuming (a washing cycle is about 100 minutes) so packaging samples to be tested were selected according to their response at the pressure test.

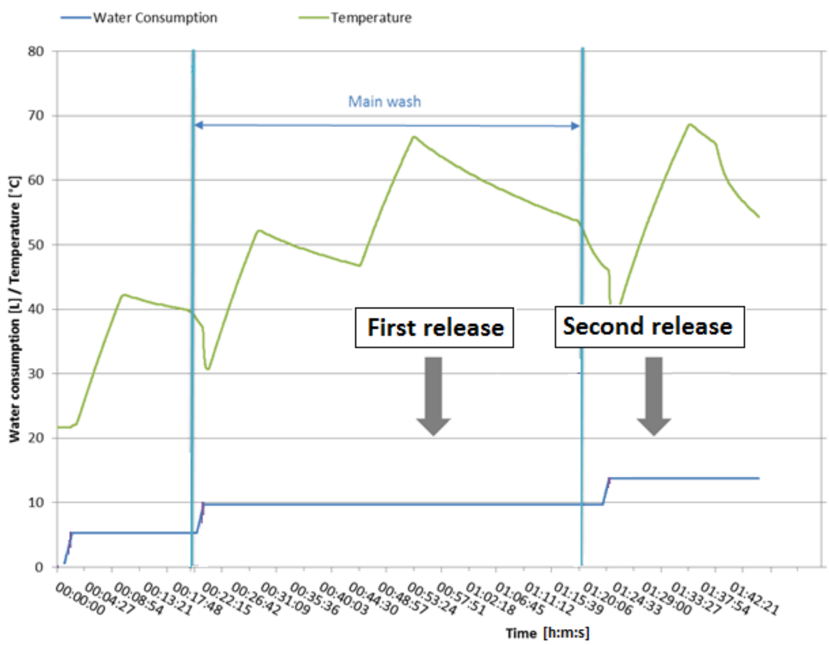


Figure 3.14: Observed time of opening of the two chambers [h:m:s], water temperature [°C], and amount of water [l] during a washing cycle.

Test of prototypes

Packaging samples were manually filled with detergent. They were tested individually in an intensive washing cycle (70 °C) without dish load, all with the same dishwasher. Sensor measured i) time of opening of the two chambers [h:m:s], ii) water temperature [°C], and iii) amount of water [l]. In Figure 3.14 there is an example of measured values during an experimental run.

The exact moment in which the two chambers open was determined. Observer can't discriminate between left and right chamber, but only between first and second release. By crossing time of opening with water temperature and amount, it can be established even in which part of washing cycle the two chambers open. Main wash last from minute 22 to minute 80. In Figure 3.14, for instance, first chamber opens during main wash, while second chamber opens during rinse that by the way is too late.

Samples produced according 8 different factors levels combinations were tested. Factors levels combination were classified according to pressure test in three groups: pressure $< 20\%$, pressure 20% - 30%, and pressure $> 30\%$.

Table 3.5: Observed opening time [minutes] and leftover liquid [%]

Classification in pressure test	Pressure test left chamber	Pressure test right chamber	Opening time 1^st release	Opening time 2^nd release	Leftover liquid left chamber	Leftover liquid right chamber
$<$ 20%	15.4	13.8	N/A	N/A	N/A	N/A
20%-30%	21.4	21.4	57	91	20	10
20%-30%	22.5	22.2	56	89	20	10
20%-30%	29.0	25.8	58	90	10	10
20%-30%	29.4	23.0	56	86	10	10
20%-30%	28.2	25.0	58	90	10	10
$>$ 30%	40.0	33.8	98	N/A	100	40
$>$ 30%	39.4	33.8	99	N/A	100	50

Data Analysis

Analysis of data aims to individuate those factors levels combinations that satisfy minimum criteria in the dimensions under investigation. Furthermore, a correlation between pressure test and opening time is investigated. Mean data observed are in Table 3.5.

In general, test reveals a basic correlation between product release and pressure:

– Pressure test value $<$ 20%: packaging does not satisfy minimum safety criteria for customer; detergent leak out during manual activation before placing in dishwasher.

– Pressure test value 20 - 30%: first detergent release is in the main wash at minute interval 56 - 58, second release is during rinse cycle; product release is in the range from 80% to 90%.

– Pressure test value $>$ 30%: only one chamber opens at the end of the rinse cycle; there is partial detergent release.

None of the selected samples reached the minimum quality criteria.

Comments on performance of final product

Dishwasher test is performed on a selection of samples from pilot line production according to pressure test results. The objective is to explore performance of final product in order to find those factors levels combinations in production that could allow respect of minimum criteria. Furthermore, a correlation between pressure test and dishwasher test results is investigated.

Contrary to pressure test, dishwasher test is time consuming. A careful selection of samples has to be done. The key success factor is the coverage of the whole range of pressure test results. Material A revealed unsatisfactory performance in the final product. Only a systematic approach allowed the investigation of the whole factorial space and the final evidence based assessment of the material. Material A was rejected.

3.4 The Experimental Strategy

The experimental strategy developed thanks to principles of Design of Experiment is composed of three phases:

1. *Material characterization.*
   Objective: Material selection.
   Success criteria: i) differentiability between weak and strong seal; ii) low variability of seal performance related to small fluctuation of production control factors; iii) low variability of seal performance related to level of control factors.
   Factors: i) Temperature of sealing jaws; ii) Time of sealing; iii) Force per cm2
   Response: Seal strength value.

2. *Production in pilot line.*
   Objective: Weak seal investigation, and selection of samples for dishwasher test.
   Success criteria: i) same behavior between left and right chamber; ii) low variability of seal performance related to levels and small fluctuation of production control factors.
   Factors: i) Pre-heating station temperature; ii) Pace of the line; iii) Time of sealing.
   Response: Burst pressure value.

3. *Performance of final product.*
   Objective: Find successful factor level combinations, and correlation between pressure test results and dishwasher test results.
   Success criteria: i) time of opening; ii) complete release of detergent.
   Response: Time of opening.

Final result is the selection of factor level combinations in production according to the desired packaging performance taking into account peculiar features of material, and correlation between factor level combination and performance. The strategy is summarized in Figure 3.15.

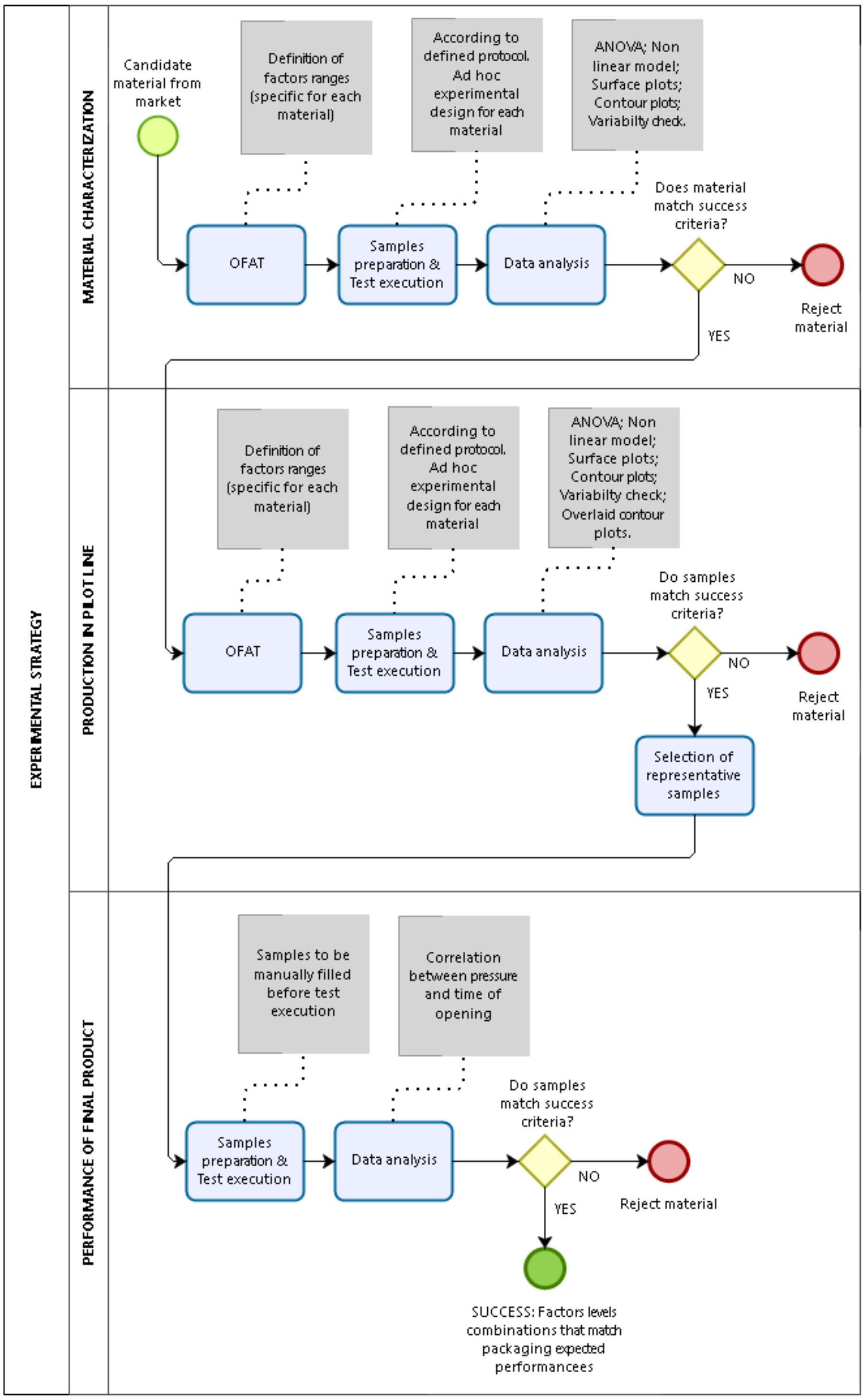


Figure 3.15: Scheme of the experimental strategy

3.5 Conclusions

An experimental strategy on innovation of thermoforming production process has been developed. Design of Experiment (DoE) techniques were used in designing and analyzing all the phases of the strategy. DoE enhanced innovation capability allowing reduction of systematic errors and distortions, full exploration of factorial space, and reduction of number of tests. The experimental strategy allows selection of material and correlation of control factors levels to packaging performance for each tested material.

Traditional approach to production control in thermoforming process was challenged. DoE allowed to identify and overcome the mismatch between control factors in laboratory and in production line. Anyway, mismatch suggests development of 2 separated sealing stations: one for strong seal and one for weak seal. In this way same control factors could be used in laboratory and in production line. Result would be a direct correlation between performance in dishwasher and control factors in laboratory. The experimental chain would be shorter and a significant reduction of number of tests should be allowed.

Chapter 4

Impact of DOE on Innovation Process Management, a Case Study

4.1 Introduction

The issue of management of innovation process is currently an important topic for many companies. As already said, innovation is one of the ways, and in some fields the most important, to achieve and maintain a competitive advantage. Innovation is a continuous process that has to be managed according to some dimensions. We have to forget the myth of a disruptive idea popping in someone's mind. Yes, it could happen, but such an unpredictable situation is not what a company could invest on. Innovation is more a cumulative process that has to be managed creating a favorable environment to let grow incremental innovation and to let breakthrough innovation be free to manifest. Innovation process is a process of recognizing customer needs and market opportunities for innovative products, generating and elaborating innovative ideas, working with knowledge regarding innovation and with information, realizing innovative activities and innovative production system to ensure a successful extension of the innovative product or service to the customers [49]. Recall that focus of my research is the innovation of production processes.

There is an increasing interest among scholars about the management of innovation process ([49], [75], [85], [60], [78], [77], [74], [22], [53], [71]). How to improve the innovation process in order to enhance company performances is investigated and dimensions that could be relevant have been identified. Some of them are definition of scope and mission, learning, communication (within organization), knowledge building, organizational structure, teamwork.

Learning is considered a critical point in the innovation management. The company should learn through the procedure in the innovation process, not only about the object of the innovation, but even concerning the process of innovation itself.

Regarding communication, an efficient information flow should be ensured in the company. If not, misunderstanding could arise and different opinion among group of people could stabilize, thus creating barriers and freezing the innovation process.

Another issue is the way information are collected and transmitted. In many cases there is no evidence of the results achieved because the documents and the records do not offer all the instruments to be properly understood and their interpretation is possible only thanks to the presence of the person that issued them. This problem is related to the issue of knowledge building inside the company. Insufficient implementation of instruments devoted to knowledge management is a problem in the field of innovation management. It could happen that the knowledge created in the innovation process is forgotten or lost. As result, part of the effort and resources in the research are wasted to repeat the creation of knowledge which has already been created previously.

The organizational structure could be a great support to the innovation process or the cause of its failure. Managers should have sufficient information about available resources and means when they plan innovative activities. The leadership and the people in charge for every activity should be clear in order to well coordinate activities and prevent independent or "hidden" research carried out only to satisfy the ego of some of the members of the team. This could lead to parallel research that, if not planned, could cause waste of time and resources.

The use of DOE for innovation is debated among scholars and practitioners as shown in section 2.2. The two opposite positions can be summarized quoting two important scholars: Montgomery [56] "Design of experiment is viewed as part of a process for enabling both breakthrough innovation and incremental innovation, without which western society will fail to be competitive" and Johnson [41] "Innovative environments thrive on useful mistakes and suffer when the demands of quality control overwhelm them. [refering to the use of DoE and Six Sigma, Ed.]". DOE is a methodology whose adoption could impact on the innovation process performances. Anyway, it's not DOE performing the innovation but people. DOE can provide a very effective and efficient aid that leads to innovation, but people are the ones that will produce results. Management of innovation process is necessary. Nevertheless, DOE could be helpful to improve the management of the innovation process providing proper instruments for managing the process.

Innovation in companies is mainly subject of the work carried out at the Research and Development department. The R&D department is an inclusive name referred to two distinct activities. The *research* activity is the one that could be motivated purely by the need to improve the knowledge in a certain field, it could be speculative and it is allowed to take risks and to fail. The *development* activity aims to realize what has been invented, it should take no risks but go straightforward to the achievement of the goal. The customer of the development activity is generally manufacturing or production process.

The development activity within a R&D department is the field where present research has been conducted. I had the opportunity to be member of a team working on the project described in details in Chapter 3. The objective was to develop a new thermoforming process to produce a new packaging releasing detergent in dishwasher at a well-defined moment during main wash.

4.2 Research Method

This research has been conducted according to the principles of qualitative research. I was part of the team whose objective was to innovate the production system in order to achieve mass production capability of the new product (the packaging for dishwasher detergent). My role was double. As expert in engineering and DOE, I was providing technical consultancy on the tools to adopt and how to use them. Namely, I was designing the fractional factorial experimental plans and analyzing data, furthermore I was involved in technical engineering issues. This part is widely described in Chapter 3. The second role was the role of *complete participant* in the adoption of DOE by the company as instrument to enhance innovation process. According to the role of complete participant described by Macri and Tagliaventi [52], I was involved in the team and worked with the team. I could observe the behavior of people involved and study the impact of the adoption of DOE on innovation process management. I was thus able to understand the opinions of the members of the team and observe their evolution in time. Weekly meetings were planned in teleconference in order to coordinate the activities. Field notes and meeting reports were the most relevant instruments to trace the evolution of the innovation process management and the impact of DOE.

4.3 Innovation Process before Adoption of DOE

The organizational structure in the innovation process was articulated in two different teams working in different R&D departments of the same multinational company situated in different countries. Furthermore, part of the process was performed at the plant of an external supplier, a company working as co-packer and potential supplier of the production line.

There was lack of a clear definition and understanding of the problem, and, above all, lack of coordination of activities. It was not clear which were the roles and which tasks should have been carried out by each team. The presence of two teams working on the same project, instead of being an opportunity, caused competition and conflicts due to subjective opinions and interpretation of facts. There was no agreement about the strategy to adopt during experimentation, and what was relevant and what was not. Pressure from marketing to achieve results put the teams in a hurry and parallel unplanned tentatives of development of the process started. Someone tried to push to invest effort on the geometry of the packaging and some others to invest effort on newly developed materials direction, but none of them had awareness of what factors had impact in production line.

The experimental strategy illustrated in Chapter 3 shows how the innovation process should have been organized in three different phases: i) characterization of materials in lab, ii) production of prototypes in the pilot production line, and iii) test of performance of the prototypes in lab. But this fact was not clear to the people involved. There were barriers due to lack of communication and proper understanding of phenomena. In Figure 4.1 the barriers in the innovation process. The barriers are between the steps of the experimental strategy and between teams and places where the innovation was carried out.

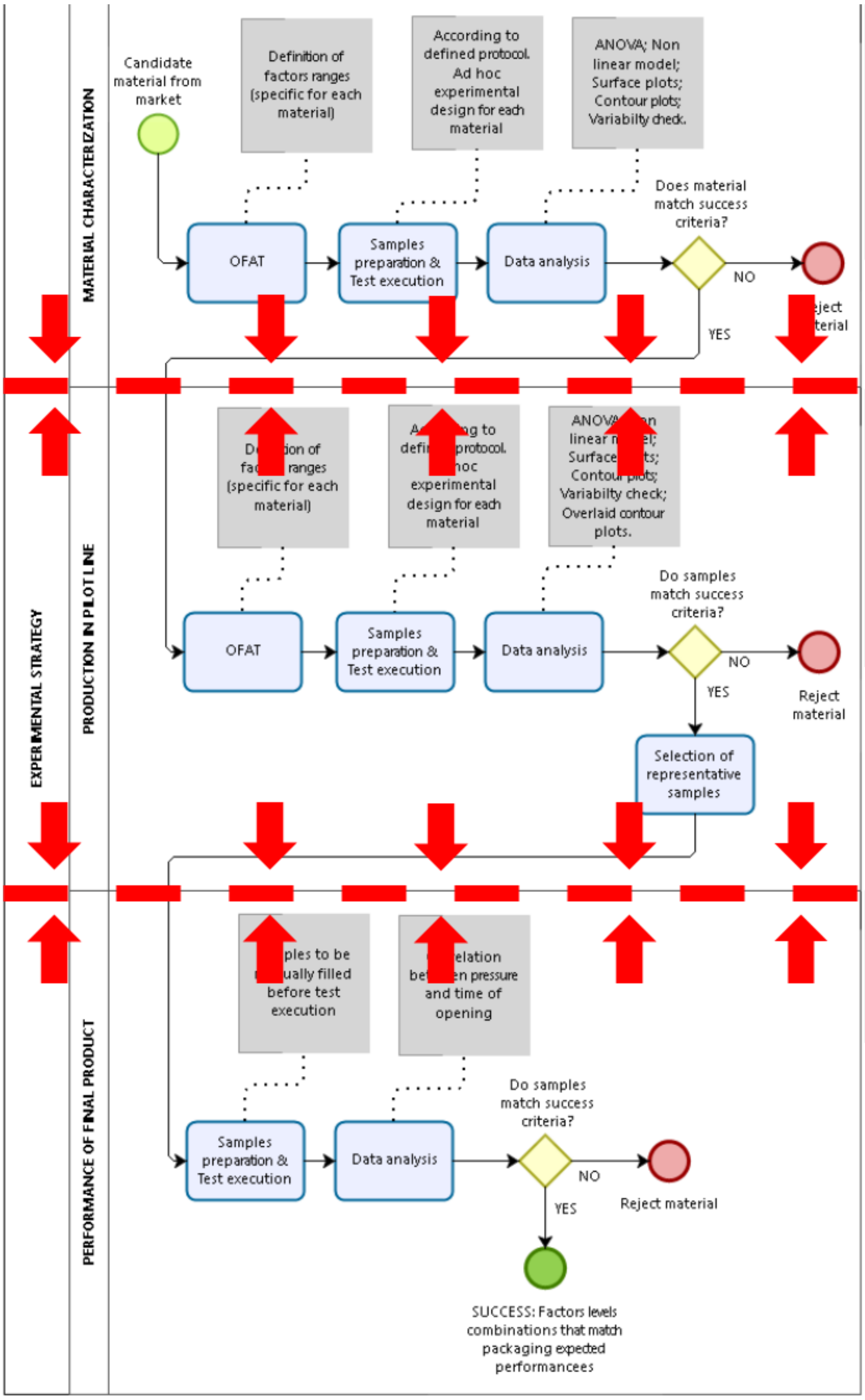


Figure 4.1: The experimental strategy developed in Chapter 3. In red the barriers impeding the innovation process.

People involved had a limited vision of the innovation project (Figure 4.2). There was not a complete understanding of the innovation process. Due to this fact, there was no correlation between results of test and control factors because of the mismatch between the control factors in lab experiments and pilot line experiments. This was a critical point that emerged thanks to the systematic experimental approach of DOE.

The production process innovation was carried out using a common theoretical modeling in packaging production. Based on that the experimental strategy was of two types quite common in the industrial environment. The first experimental strategy was the one-factor-at-a-time (OFAT) strategy. This strategy was used in the characterization of materials phase. It consists in holding constant all the factors that are supposed to have an impact on the result a part from one that varied over its range. The main problem of this procedure is that it doesn't detect the possible interactions between factors. The second experimental strategy was the best-guess approach. In this case the experimenter makes a first guess based on his or her experience and engineering knowledge. Then, based on the outcome of the first experiment, another experiment is planned. This approach is carried out until the success is achieved, or until experimenters get discouraged by results and abandon the effort. This strategy was used in the production plant and the experimenters were stuck in a trial-and-error vicious circle.

A side effect of the beliefs and behavior of some members of the team was the low level of quality during the experiments. The lack of understanding the relevance of control of variability and the importance of repeatability didn't allow to draw meaningful conclusion from dataset collected. Charles Hicks, a famous professor of statistics at Purdue University, is said to be used to teach that "...if you have 10 weeks to solve a problem, you should spend 8 weeks planning the experiment, one week running it, and one week analyzing the data" [56]. Recall that all experiments are designed experiments, but the way you design them makes the difference. A poorly planed experimental design will usually lead to disappointing results, while a carefully planned one will usually produce helpful results.

The project was lunched one year earlier than my team was involved. They already had performed 20 days of experiment at the production plant. Results collected during experiments were recorded in excel files with poor notes on the meaning of different variables and on experimental settings. It happened to receive some old data set to be analyzed in order to have a first idea about variability of responses, but interpretation of variables meaning was not clear even to the people that was providing the data set. Some fundamental information were basically retrieved thanks to the memory of some of the participants.

The data were provided by excel files with just some basic synthesis variable without further analysis. Results interpretation was not univocal among the people involved. Furthermore, most of the times people didn't make any effort to interpret a data-sheet with a lot of columns and rows filled by numbers that couldn't show any clear information just visualizing them.

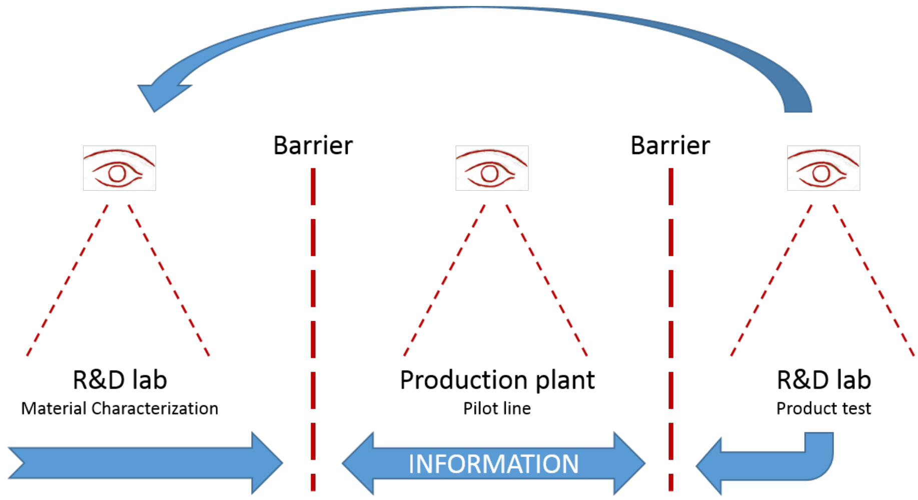


Figure 4.2: Sketch of the organization according to information flow and vision on the process. People (the eyes) involved on the innovation process had a limited vision of the process. The information flow (blue arrows) was hindered.

The most relevant observations concerning the innovation process management before the adoption of DOE are:

- Information was not structured and well codified, it was shared on demand and difficult to interpret.
- There was lack of cooperation. Competition was due to subjective opinions and interpretation of facts.
- The process was stuck in a trial and error approach and there was no progress.
- The innovation project was launched one year earlier. They were already performed 20 days of experiments at the pilot line.
- A mismatch between control factors in lab experiments and pilot line experiments was not discovered.

4.4 The Adoption of DOE

In this section I describe some of the most relevant steps of the adoption of DOE from the point of view of the innovation process management.

The introduction of a new support to innovation was not easy since the beginning. This fact can be easily understood within the topic of *resistance to change* that is well known in the managerial field. Generally speaking, people can be reluctant to modify their behavior because they think that they are doing a good job, that they don't need any external interference, and because they want to defend their good reputation. Therefore, the presence of an important sponsor was fundamental to introduce a change in an environment skeptical towards the new methodology proposed.

First thing was to move a few steps back from the experimental phase and start back from the scratch. DOE according to Montgomery [54] moves first steps starting from i) recognition of and statement of the problem, ii) choice of factors, levels, and ranges, iii) selection of response variables. This was done both at lab level and at production plant level. The intent was to clarify the purpose of the research and define problems and goals clearly. A common understanding of the objective of the experimentation, of the strategy and of the expected results was the beginning of a new attitude and a more deep involvement of all the members of the team. The innovation process can be greatly catalyzed by group discussion if the group contains people from different disciplines.

Providing a holistic view and system thinking allows to see how the different parts of the system interact. We wanted to change the perspective from which the problem was tackled by asking fundamental questions and challenging basic assumptions. End to end involvement on the project was fundamental to discover barriers between teams of experimenters and misleading subjective perspectives on the problem. It was the understanding of the complete process that allowed innovation process to be most effective.

At the beginning DOE was viewed as more time-consuming and difficult than traditional approaches. The fractional factorial design is less intuitive compared to other methods such as OFAT. Furthermore the team was under pressure because of the deadlines set by marketing department. This caused a rush to find quick solutions to fix immediate problems. The solutions adopted revealed to be ineffective, and the consequence was a further delay and waste of time and resources. The members of the teams had a weak statistical background that inhibited their understanding and use of DOE. Indeed, DOE was seen by some members of the team as an academic exercise not leading to concrete and practical results.

The awareness of the relevance of control of variability and the importance of repeatability was raised. Data collection has to be performed in the proper way. The team had to be sure that they are sampling appropriately. Experiments have to be reproducible, so an experimental protocol had to be defined. The variability issue that experimenter have had to face before the introduction of DOE, and whose causes where searched by analyzing data, was caused basically by an incorrect procedure during experiments execution.

The building of a series of experiments should be adapted basing on knowledge gained from earlier experiments and refined over time. How can it be possible if previous learning are not codified in such way that results could be retrieved and reused? A fundamental aspect of use of DOE is the sequential knowledge gain that comes from use of scientific method and from a proper way of coding experiments and results.

DOE approach facilitate decision-making based on quantitative objective information instead of subjective opinions. Indeed, DOE revealed to be useful as a common language that tied together people from different teams and diverse disciplines. The results of the analysis of data collected during experiments were shown to the team during meeting thanks to slide presentation where all the information needed were included. Namely: the specific objective of the experiment, the variables involved, the settings, the experimental design, graphical analysis, visual representation of the analysis (e.g. response surface), conclusions, and open points to be further investigated.

The initial skepticism about DOE was defeated even thanks to the capability to achieve significant results by means of a very reduced number of experiments, and thanks to showing results of the analysis in a more effective and intuitive way by response surfaces and contour plots. Compared to previous excel files used to communicate results, the graphical methods highly enhanced the communication and the understanding among the experimenters. Two examples are given in Figures 4.3 and 4.4. The response surface in Figure 4.3 allows to understand why in certain ranges of values of control factors the response is more stable (flat area) compared to other ranges (steep slope area) thus allowing an higher control. Indeed, small variation in control factors has almost null effect on response variable in the flat area. This graphical representation gave information whose quality and effectiveness were far above an excel data sheet. The contour plot in Figure 4.4 had a similar effect on participants. Indeed, it gave the direction to improve the response setting the control factors in production. Generally speaking, graphical representation of experimental results allows to discover patterns and information in the data. It improves communication and the achievement of a shared opinion about a specific issue. Looking at the same data in a different way can lead to a better understanding of the impact that control factors could have on the response variable, thus clarifying correlations and cause-and-effect relationships.

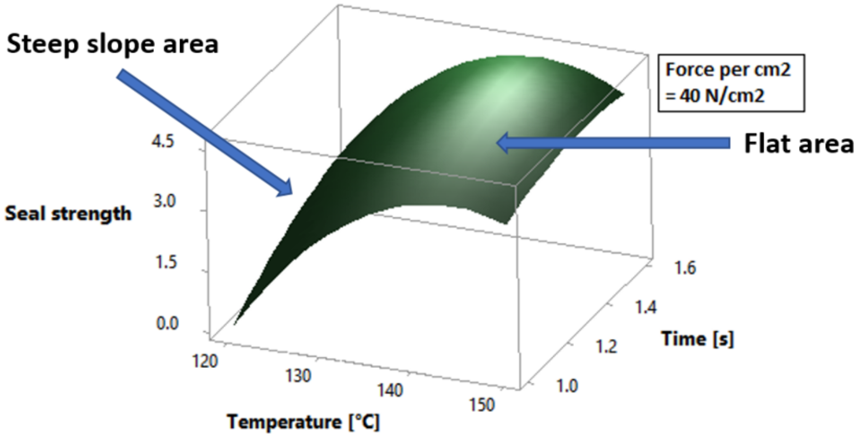


Figure 4.3: Example of response surface.

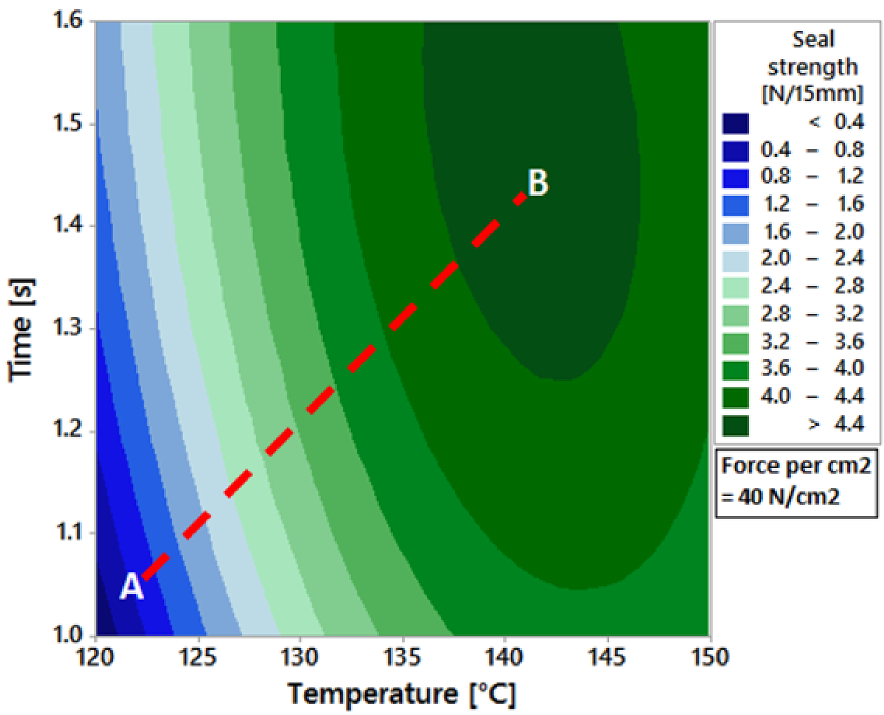


Figure 4.4: Example of contour plot.

An important result was achieved when it was shown the mismatch between control factors in lab and in pilot line. This fact strengthened the need to have a common goal and to work as one team. The members of the team felt more involved and enthusiastic. For the first time, a correlation among all steps of the innovation process was achieved.

The multivariate analysis was required but never used before the adoption of DOE. Presenting different views of the data using overlaid contour plots (Figure 4.5) encouraged a different approach that differs from the conventional view of the system. Again, changing the perspective from which the problem was tackled was a key to success.

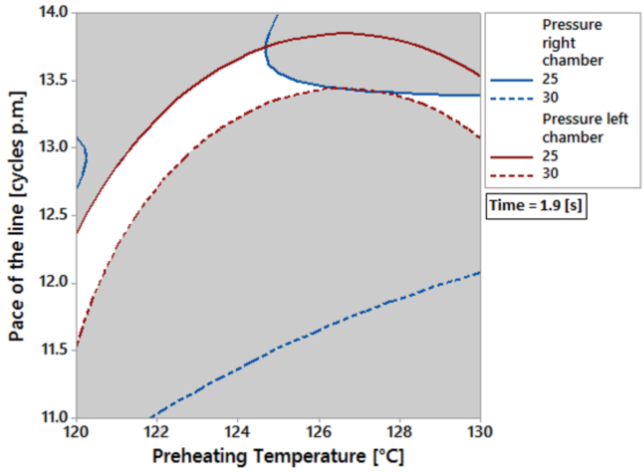


Figure 4.5: Example of overlaid contour plot.

4.5 Innovation Process after Adoption of DOE

The introduction of DOE brought a new perspective on the innovation process. The new perspective was achieved even by asking fundamental questions and challenging basic assumptions. But, the most important thing was that it was a common perspective among the members of the team (Figure 4.6). The vision of the process became broad and there were a clear interpretation of correlation between test results and control factors. The latter was not possible before the adoption of DOE. The common goal and the common understanding of the objective of the experimentation lead people to work as one team. The barriers between the teams of experimenters were removed thanks to the end to end involvement.

DOE became a common language. It facilitated decision making by the team avoiding conflicts due to subjective opinions. The way results of analysis are shown enhances communication. The quality and effectiveness of information were increased. This creates a fertile field to knowledge building, by means even of the improved system of coding experiments and recording results.

There is anyway one risk that I perceived. The risk is that someone performs unnecessary tests only to seek for the "DOE blessing" in order to not expose himself to criticism. This behavior is contrary to the reason why DOE should be adopted and the benefits it can give. The experimental design should always move first steps from the definition of the problem and not jumping to the experimental phase just to show some data analysis.

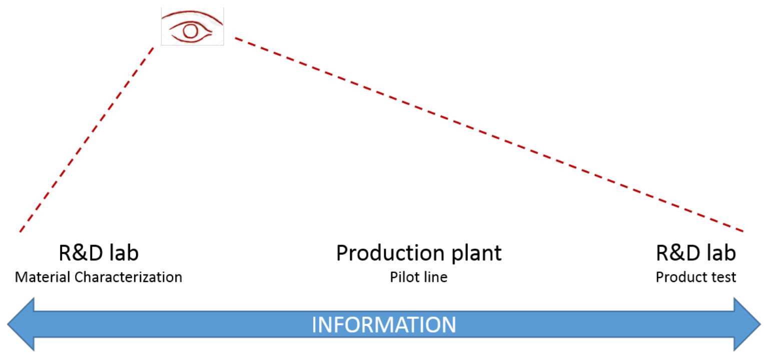


Figure 4.6: Sketch of the organization according to information flow (blue arrows) and vision on the process. People involved (the eyes) on the innovation process now have a broad vision of the process. The information flow is enhanced.

The most relevant observations concerning the innovation process management after the adoption of DOE are listed below. They can be compared to the same observations made about the innovation process management before the adoption of DOE.

- DOE became a common field and a common language for information exchange and knowledge building.
- There is full cooperation among the members of the team based on objective and quantitative data interpretation.
- An exploration of the whole factorial space was accomplished, thus enabling progress in innovation.
- Team weekly meetings were scheduled from August to October. One week of experiments in lab was performed, then 3 days at pilot line, then one week in lab again to achieve the result.

- The mismatch between control factors in lab experiments and pilot line experiments was overcomed.

4.6 Suggested benefits of Adoption of DOE

The qualitative research performed during the adoption of DOE for the development of a new packaging production process allowed to highlight the impact of DOE on innovation process management described on previous sections and specific to the case analyzed. In this section the aim is to generalize the learnings beyond the specific case along five dimensions that have a more general value in the managerial field. Namely: decision making, integration, communication, time and cost, and knowledge.

The benefits suggested by case study are listed below:

- Decision making

  DOE allows decision making to be objective and quantitative instead of being based on subjective opinions.

  DOE forces to ask the right questions to clarify the purpose and the issues. It helps to define the goals, hypotheses and problem clearly.

  DOE does not focus only on data analysis perspective, but also on data collection perspective.

  DOE helps to switch from a narrow channel of thought (best guess approach or OFAT) to broader vision and range of information (Full or fractional factorial space). Designed experiments accelerate learning and evaluations of control effects, they streamline decision making about how to proceed.

  Presenting different views of the data (numerical and graphical summaries) encourages thinking that differs from the conventional view of the system. DOE helps to display and analyze processes, variation and data in a way that cause-and-effect relations are clarified and process understanding is enhanced.

- Integration

  A broad vision of innovation process is enhanced thus allowing the end to end involvement of all the people of the team.

- Communication

  DOE can be the common language that ties together collaborators from diverse disciplines.

- Time and costs

  DOE allows a strong reduction of time and number of experiments.

- Knowledge

  DOE helps to achieve a complete understanding of the process and knowledge of inputs and variation.

  The way data are collected, organized, analyzed and recorded allows data retrieval and re-usability.

Some of these benefits are supported by scholars in literature. Nevertheless, they should be further investigated through more case studies.

Chapter 5

Recent Nonparametric Methods Compared and Contrasted: a Simulation Investigation in Factorial Designs

5.1 Introduction

Industrial experiments are commonly based on factorial designs. Design of Experiment (DOE) is popular in different fields of engineering as for instance for bio-fuel production [83], for industrial production practices [38, 79], for machines' production process [44], for alternative raw materials experimentation [43], or for the use of Coordinate Measuring Machines for quality control [72]. The two-way two-levels crossed factorial design is a common design used in the exploratory phase. Thanks to two-way multi-level design, practitioners can assess the impact on response variables of two factors they can control during the experiment and of their interaction according to the assumed model. From here on, we designate the 2 factors as factor A and factor B, and their corresponding levels as (A_1,A_2) and (B_1,B_2), respectively.

The F test in the usual linear model for analysis of variance (ANOVA) is a common instrument to compare the means of observations grouped according to factor level combinations, but sometimes data sets do not satisfy the assumptions of parametric tests. When assumptions like normal distributions of errors and homoscedasticity are violated, nonparametric tests are powerful instruments to support data-based decisions [36].

A variety of nonparametric methods have been developed during recent years. There are permutation tests and rank-based tests available for practitioners in scientific literature which are implemented in R software packages or functions. Furthermore, there are robust alternatives [80] as well as approximations for parametric tests recommended in case of unequal variances and small sample sizes [11]. In general, permutation tests are computationally intensive and distribution free. In this study we focus on Constrained Synchronized Permutations (CSP) [3], Unconstrained Synchronized Permutations (USP) [3], Wald Type Permutation (WTP) [61], Aligned Rank Transform (ART) [31] and ANOVA-Type Statistic (ATS) [11],which are designed to address the same hypotheses.

Practitioners have to convert the objective of the experiment in terms of type I and type II error rates. According to the objective, different levels of power of the test are required. Expected power of a test should be taken into account starting from the design phase of the experiment. Power of common parametric tests has been widely investigated and many software implement functions to calculate the power of a test according to the data set to be analyzed and other parameters (Faul et al. 2007: G*power 3; pwr package in R). It is well known that factors that impact the power of a parametric test are: α level, variance, factor effect (expected difference between means) and number of replicates. For instance, in the design phase of the experiment, practitioners could estimate expected variance of data from historical data, they could define factor effect they want to investigate according to the objective of the experiment, and finally they could define minimum number of replicates according to the needed power. Number of replicates could be a strong constraint when experiment is expensive.

Performances of nonparametric methods compared in this study have been assessed in recent publications with respect to the nominal α level. In particular, some data are available on the performances of CSP [25, 3], USP [25, 3], WTP [61, 25], ATS [61, 11, 25], ART [31]. While power of the test has been investigated for CSP [25, 3], USP [25, 3], WTP [61, 25, 26], ATS [61, 11, 25], ART [31]. To the best of our knowledge, there is no comparison of power of the aforementioned nonparametric tests under the same conditions. In fact, simulation designs in previous studies vary significantly, and a fair comparison between the five methods based on existing results is not possible. Furthermore, the possible impact that the level of a non-investigated factor could have on the power of the test on a main factor has not been considered for all the methods in the two-way designs.

In this study we primarily compare the power of CSP, USP, ART, WTP, ATS and F tests in a two-way two-levels balanced factorial design. We fix the level α at 0.5 and we assess the power ($\mathrm{P_w}$) along three dimensions of data set: i) factor effect, ii) standard deviation, and iii) number of replicates. The objective is to assess the impact of the three dimensions on the power of the tests and to give useful information on the choice of the test. We consider both the homoscedastic and the heteroscedastic cases. Furthermore, in a smaller scale simulation, we investigate the performances of the test in a two-way two-level unbalanced design varying the factor effect both in the homoscedastic and heteroscedastic case.

Some interesting findings are related to (i) the different sensitivity of the tests to the variation of the dimensions, (ii) the conditions under which some tests can't be used, (iii) the tradeoff between power and type I error, and (iv) the bias of the power on one main factor analysis due to presence of effect of the other factor.

5.2 The linear additive model and the test hypotheses

First, the notation used in the description of the investigated methods is herein illustrated. We assume the linear additive model of a two-way crossed factorial design:

$$Y_{ijk} = \mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}+\epsilon_{ijk}, \quad i=1,2; \quad j=1,2; \quad k=1,\dots,n_{ij}, \tag{5.1}$$

where μ is the overall mean, α_i is the effect of level i of factor A, β_j is the effect of level j of factor B, $(\alpha\beta)_{ij}$ is the effect of interaction between factor A at level i and factor B at level j, and ϵ_{ijk} is the error term. The mean of error term is $E(\epsilon_{ijk}) = 0$ for each factor level combination, and n_{ij} is the number of replicates. The total number of observations is $N = \sum_i \sum_j n_{ij}$

The side conditions are:

$$\sum_i \alpha_i = 0; \quad \sum_j \beta_j = 0; \quad \sum_i (\alpha\beta)_{ij} = 0 \ \forall j; \quad \sum_j (\alpha\beta)_{ij} = 0 \ \forall i. \tag{5.2}$$

The null hypotheses of no-main effect of factor A, no-main effect of factor B and no-interaction effect between factors A and B are:

$$\begin{aligned} H_0^{(A)} &: \quad \alpha_1 = \alpha_2 \\ H_0^{(B)} &: \quad \beta_1 = \beta_2 \quad \text{and} \\ H_0^{(AB)} &: \quad (\alpha\beta)_{11} = (\alpha\beta)_{12} = (\alpha\beta)_{21} = (\alpha\beta)_{22}, \end{aligned} \tag{5.3}$$

respectively. In vector notation, the data can be written as $\boldsymbol{Y} = (Y_{111}, \dots, Y_{11n_{11}}, Y_{121}, \dots, Y_{22n_{22}})'$. And the null hypotheses can be written in terms of contrasts as:

$$\begin{aligned} H_0^{(A)} &: \quad \boldsymbol{C}_A\boldsymbol{\mu} = \boldsymbol{0}, \\ H_0^{(B)} &: \quad \boldsymbol{C}_B\boldsymbol{\mu} = \boldsymbol{0} \quad \text{and} \\ H_0^{(AB)} &: \quad \boldsymbol{C}_{AB}\boldsymbol{\mu} = \boldsymbol{0}, \end{aligned} \tag{5.4}$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})^{'}$, $\mu_{ij} = E(Y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ and $\boldsymbol{C}_M$ is a full row rank contrast matrix for $M \in \{A, B, AB\}$.

5.3 A Parametric Method

The ANOVA-type statistic (ATS) was introduced by Brunner et al. [11]. They proposed a method for approximating the distribution of quadratic forms with the aim of improving the accuracy of testing hypotheses in linear models with heteroscedastic error structure and small sample sizes.

Considering the null hypotheses formulated in terms of contrast matrices $\boldsymbol{C}_M$ in section **??** for $M \in \{A, B, AB\}$, let $\boldsymbol{T}_M = \boldsymbol{C}_M^{'}(\boldsymbol{C}_M\boldsymbol{C}_M^{'})^{-1}\boldsymbol{C}_M$ be the orthogonal projection. It can be shown that $\boldsymbol{C}_M\boldsymbol{\mu} = \boldsymbol{0}$ if and only if $\boldsymbol{T}_M\boldsymbol{\mu} = \boldsymbol{0}$. As consequence, the null hypotheses of no-main effects and of no-interaction can be written as:

$$\begin{aligned} H_0^{(A)} &: \quad \boldsymbol{T}_A\boldsymbol{\mu} = \boldsymbol{0}, \\ H_0^{(B)} &: \quad \boldsymbol{T}_B\boldsymbol{\mu} = \boldsymbol{0} \quad \text{and} \\ H_0^{(AB)} &: \quad \boldsymbol{T}_{AB}\boldsymbol{\mu} = \boldsymbol{0}. \end{aligned} \tag{5.5}$$

The ATS is defined by:

$$F(M) = \frac{N\overline{\boldsymbol{X}}^{'}\boldsymbol{T}_M\overline{\boldsymbol{X}}}{\text{trace}(\boldsymbol{T}_M\boldsymbol{S})} \tag{5.6}$$

The null distribution of $F(M)$ is approximated by F-distribution with degrees of freedom:

$$f_1 = \frac{[trace(\boldsymbol{T}_M\boldsymbol{S})]^2}{trace[(\boldsymbol{T}_M\boldsymbol{S})^2]} \qquad and \qquad f_2 = \frac{[trace(\boldsymbol{T}_M\boldsymbol{S})]^2}{trace(\boldsymbol{D}_M^2\boldsymbol{S}^2\boldsymbol{\Lambda})}, \tag{5.7}$$

where $\boldsymbol{D}_M$ is the diagonal matrix of the diagonal elements of $\boldsymbol{T}_M$, $\boldsymbol{S} = N diag\{\hat{\sigma}_{ij}^2/n_{ij}\}_{i,j=1,2}$, $\hat{\sigma}_{ij}$ is the empirical variance in cell (i,j), $N = \sum_i \sum_j n_{ij}$ and $\boldsymbol{\Lambda} = diag\{(n_{ij}-1)^{-1}\}_{i,j=1,2}$.

The ATS relies on the assumption of normally distributed error terms and it is an approximate test. According to Brunner et al. [11], in a two-way or higher-way design with normally distributed errors the ATS has to be preferred to the classical ANOVA because it maintains the nominal level with high accuracy and it is robust under heteroscedasticity. Furthermore, in the homoscedastic case it has the same performances in terms of size of the test and power as classical ANOVA–F test [11].

5.4 A Rank-Based Method

The Aligned Rank Transform (ART) [32] is a rank-based nonparametric method for factorial designs. It is a five step procedure in which data are aligned, ranked and analyzed using an appropriate parametric procedure. In the case of this study, the two-by-two design, the common ANOVA based on F-test will be used. ART has been introduced by Higgins et al. [31] with the aim of overcoming the limitations of rank transform procedure [14] in detecting the interaction effect in a factorial analysis. The rank transform flaw is due to the fact that the interaction structures that exist in the original data may not anymore be present in the dataset after a non-linear transformation such as the rank transform [31]. That's why the alignement has been introduced.

Assuming a linear additive cell mean model, the alignment procedure consists in isolating the investigated factor effect in the response variable by "removing" the non investigated factor effects. This procedure allows to achieve a great improvement in performances in detecting interaction effect [31]. We herein briefly illustrate the ART steps for the two-by-two design:

Step 1 Compute residuals. Residuals are the difference between the observed values and the mean of the values of the factor level combination they belong to.

$$r_{ijk} = Y_{ijk} - \overline{Y}_{ij} \tag{5.8}$$

for $i = 1, 2$; $j = 1, 2$ and $k = 1, \ldots, n_{ij}$ where $\overline{Y}_{ij} = \sum_{k=1}^{n_{ij}} Y_{ijk}/n_{ij}$.

Step 2 Compute estimated effect for factor A, factor B and for interaction.

$$\begin{aligned}
\hat{\alpha}_i &= \overline{Y}_i - \hat{\mu} \qquad i = 1, 2 \\
\hat{\beta}_j &= \overline{Y}_j - \hat{\mu} \qquad j = 1, 2 \\
(\hat{\alpha\beta})_{ij} &= \overline{Y}_{ij} - \overline{Y}_i - \overline{Y}_j + \hat{\mu} \qquad i = 1, 2 \qquad j = 1, 2 \quad ,
\end{aligned} \tag{5.9}$$

where

$$\overline{Y}_i = \left(\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}\right) / \left(\sum_{j=1}^{2} n_{ij}\right), \quad \overline{Y}_j = \left(\sum_{i=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}\right) / \left(\sum_{i=1}^{2} n_{ij}\right) \quad \text{and}$$
$$\hat{\mu} = \left(\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n_{ij}} Y_{ijk}\right) / \left(\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}\right) \tag{5.10}$$

.

Step 3 Compute aligned response for factor A, factor B and for interaction.

$$\begin{aligned} Y_{ijk}^{(A)} &= r_{ijk} + \hat{\alpha}_i \\ Y_{ijk}^{(B)} &= r_{ijk} + \hat{\beta}_i \\ Y_{ijk}^{(AB)} &= r_{ijk} + (\hat{\alpha\beta})_i \end{aligned} \tag{5.11}$$

for $i = 1, 2$, $j = 1, 2$ and $k = 1, \ldots, n_{ij}$. The aligned responses for each factor take into account only the contribution of the factor of interest and of the residuals assuming a linear additive cell mean model.

Step 4 Assign averaged ranks. Each of the three groups of aligned responses $\boldsymbol{Y}^{(M)} = \left(Y_{111}^{(M)}, \ldots, Y_{11n_{11}}^{(M)}, Y_{121}^{(M)}, \ldots, Y_{22n_{22}}^{(M)}\right)'$ is ranked, thus obtaining the vector of ranks $\boldsymbol{R}^{(M)}$ where $M \in \{A, B, AB\}$. In case of ties among l values, the average rank is the sum of ranks divided by l.

Step 5 Perform full factorial ANOVA. Three separate full factorial ANOVA (i.e. including factor A, factor B and interaction) are performed on the three aligned ranked dataset. Only the results referred to the effect for which data were aligned will be considered. That is, p-value of factor A is computed on $\boldsymbol{R}^{(A)}$, p-value of factor B on $\boldsymbol{R}^{(B)}$ and p-value of interaction on $\boldsymbol{R}^{(AB)}$.

The ART procedure has the advantage that it does not require normally distributed data and response may be continuous or discrete. ART shows limitations in case of very high proportions of ties in the original data, or in case of extremely skewed distribution as the skewness is reduced by the rank transformation [81].

5.5 Constrained and Unconstrained Synchronized Permutation Tests

Synchronized permutation is obtained according to two basic concepts. The first is that values can be permuted between two levels of a factor while keeping the level of remaining factors in the model constant. For instance, to test the significance of main effect A in the two-way design, observations can be exchanged between groups with factor A level 1 and 2 that have level of factor B equal to 1, and separately between groups with factor A level 1 and 2 that have level of factor B equal to 2. The second basic concept in Synchronized permutation is exchanging the same number of units within each pair of the considered groups [3]. Synchronized permutation assume that error terms are exchangeable that may not hold in case of heteroscedastic error variance.

According to Basso et al. [3] the test statistics for the main factors and the interaction effect in the two-way two-levels design are:

$$\begin{aligned} T_A &= (T_{11} + T_{12} - T_{21} - T_{22})^2, \\ T_B &= (T_{11} + T_{21} - T_{12} - T_{22})^2 \quad \text{and} \\ T_{AB} &= (T_{11} - T_{12} - T_{21} + T_{22})^2, \end{aligned} \tag{5.12}$$

where, for each permutation,

$$T_{ij} = \sum_k Y_{ijk}. \tag{5.13}$$

Then p-value is calculated as the proportion of permutations for which test statistics values of permuted data sets are greater or equal to the test statistic value for the original data set.

There are two ways to obtain a synchronized permutation, namely Constrained Synchronized Permutation (CSP) [69] and Unconstrained Synchronized Permutation (USP) [69].

5.5.1 Constrained Synchronized Permutation (CSP)

CSP consists of applying same permutation in all couples of groups given the initial order of observations. For instance, in the two way design if a permutation consists in exchanging the second observation of group A_1B_1 with the first observation of group A_2B_1 when testing main effect A, then same permutation has to be applied to groups A_1B_2 and A_2B_2. Observations should be randomized within each group at the beginning before performing the permutation test.

As a result of the use of the same permutation between all possible pairs of groups, the number of possible ways to exchange units is correlated only to number of replicates n in the balanced design. If n is too small, CSP could give a minimum achieved significance level higher than the desired type I error. In particular the minimum achievable significance error is $\alpha_{\min} = 2 \times (C_{csp})^{-1}$ where C_{csp} is the total number of possible permutations of CSPs, i.e.

$$C_{csp} = \binom{2n}{n}. \tag{5.14}$$

5.5.2 Unconstrained Synchronized Permutation (USP)

USP, unlike CSP, can apply different permutations in the various pairs of groups. However, the basic principle of synchronized permutations of exchanging the same number of observations has to be met. The algorithm provided in Basso et al. [3] guarantees the values of the test statistic to be equally likely. This procedure allows to overcome those cases in which the test statistic is not uniformly distributed. However, USP is computationally more intensive compared to CSP, and it is recommended in the case of small number of replicates.

5.6 Permutation of Wald-Type Statistics

The Wald-Type Permutation (WTP) has been developed by Pauly et al. [61] by applying a permutation technique to the Wald Type Statistic (WTS). The WTS is asymptotically exact in general factorial design for N $\to \infty$ even in the case of heteroscedastic and nonnormal errors. The problem with WTS is that the rate of convergence is rather slow and, for small sample size, it does not maintain the type I error, resulting in liberal tests [25]. WTP improves the performance of WTS in case of small sample size. WTP has broader scope of applicability than most permutation methods (e.g.,[62, 19, 23]) which require exchangeability under the null hypothesis.

The permutation procedure is herein briefly summarized for the two-way two-levels factorial design. Given a random permutation of all N components of data $\boldsymbol{Y}^* = (Y^*_{111}, \dots, Y^*_{11n_{11}}, Y^*_{121}, \dots, Y^*_{22n_{22}})'$, the vector of means under this permutation, denoted by $\overline{\boldsymbol{Y}}^* = (\overline{Y}^*_{11}, \overline{Y}^*_{12}, \overline{Y}^*_{21}, \overline{Y}^*_{22})'$, is calculated, where $\overline{Y}^*_{ij} = n^{-1} \sum_{k=1}^{n} Y^*_{ijk}$. The empirical variance of permuted observations is $\hat{\sigma}^{2*}_{ij} = (n-1)^{-1} \sum_{k=1}^{n} (Y^*_{ijk} - \overline{Y}^*_{ij})^2$, and $\hat{\boldsymbol{S}}^*_N = \mathrm{diag}(\hat{\sigma}^{2*}_{11}, \hat{\sigma}^{2*}_{12}, \hat{\sigma}^{2*}_{21}, \hat{\sigma}^{2*}_{22})$ is a diagonal matrix. Finally, the permuted Wald type statistic is:

$$W^*_N(M) = N(\overline{\boldsymbol{Y}}^*)' \boldsymbol{C}'_M (\boldsymbol{C}_M \hat{\boldsymbol{S}}^*_N \boldsymbol{C}'_M)^+ \boldsymbol{C}_M \overline{\boldsymbol{Y}}^*. \tag{5.15}$$

Then p-value is calculated as the ratio between the number of test statistics of permuted data set as or more extreme than the test statistic of the original data set and the total number of permutations.

WTP is applicable to general factorial designs without the assumption of homoscedasticity or normal distribution for the errors. It is not restricted to two-way factorial designs. It is applicable in higher-way classifications, in nested designs and in unbalanced designs as well. Furthermore, it is asymptotically exact when data are not exchangeable as in the heteroscedastic case, in the sense that it maintains the preassigned type I error for large sample size.

Lastly, according to Pauly et al. [61], the behavior of the test depends only on the investigated effect. In other words, the result of the test on one main effect is not affected by the fact that the null hypotheses on the remaining main effects and interactions are true or not. For instance, in the two-by-two design, the test on the factor A should not be affected by whether the null hypotheses on factor B and interaction AB are true or not.

5.7 The Simulations Campaign

A simulation study is performed to assess the power of the tests described in section **??**. Data set for simulations are generated according to the cell mean model of a two-way crossed factorial design (factor A and factor B),i.e.

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}. \tag{5.16}$$

where $i = 1, 2$ is the level of factor A; $j = 1, 2$ is the level of factor B; and $k = 1, \ldots, n$ is the k^{th} replicate for each factor level combination. In this setup, the general mean $\mu = 0$, the interaction is given by the product of effects of the two factors, and the number of replicates n is constant for all the factor level combinations.

Four distributions are used to generate the error term ϵ. Three are symmetric: normal, Laplace and student's t. One is skewed: lognormal. Both homoscedastic and heteroscedastic cases are considered. Scheme of the heteroscedasticity for the various factor levels combinations is: $1{\times}\sigma$ (A_1B_1), $1.5{\times}\sigma$ (A_1B_2), $1.5{\times}\sigma$ (A_2B_1), $2{\times}\sigma$ (A_2B_2). Here, σ refers to the standard deviation or to the scale parameter of the distribution used to generate the error.

The power of the tests will be investigated along three dimensions: i) effect of factors (α_i and β_j), ii) standard deviation (σ), and iii) number of replicates (n). Six combinations of the three dimensions of investigations have been considered as described in table 5.1. In each simulation study, only one dimension is allowed to vary at a time, while value of the remaining two are fixed. For each of the six simulations, data set are generated from the four distributions above in the homoscedastic as well as heteroscedastic cases.

Effect of factors are varied in a range from 0 to 1 in steps of 0.1. Effect of factor is the difference (δ) between the means of two different levels of a factor. As the model has two factors, all the possible combinations of effect of factor A and effect of factor B will be considered. The parameter σ is a measure of the variability of data in each factor level combination. Its effect is investigated by varying it from 0.1 to 1 in steps of 0.1, and the investigation is limited to the normal, Laplace and lognormal distributions. Number of replicates is investigated in a range from 5 to 25 in steps of 5. Small number of replicates is considered in order to accomodate the needs of practitioners who may need to design effective experimental plans in situations where there are constraints or limited resources for the experimental design. As the factorial design is balanced, number of replicates refers to each factor level combination. Furthermore, the behavior of tests in case of heavy-tailed distribution will be investigated using the student's t distribution with 3 degrees of freedom to generate error terms.

All simulations are performed in R (version 3.4.0; R Development Core Team (2017)). The number of simulations is $n_{sim} = 10000$. Number of permutations for CSP, USP and WTP is $n_{perm} = 2000$. ATS test and WTP test are performed using package GFD [20, 21], CSP and USP tests are performed using functions provided by Basso et al. [3] and ART test is performed using package ARTool [42].

Table 5.1: Simulations design. In each simulation (row) two dimensions are fixed and one is investigated.

	Investigated dimensions		
Simulation	Factor effect	σ	Number of replicates
Setting 1	0 to 1 by 0.1	1	5
Setting 2	0 to 1 by 0.1	0.5	5
Setting 3	0 to 1 by 0.1	0.25	5
Setting 4	1	0.1 to 1 by 0.1	5
Setting 5	1	1	5 to 25 by 5
Setting 6	1	0.5	5 to 25 by 5

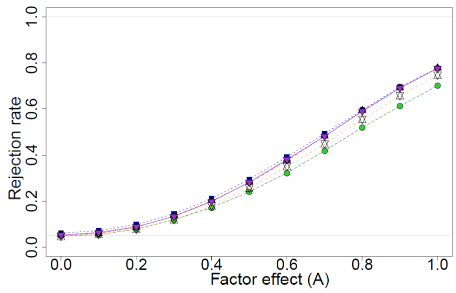
5.7.1 Results

In this section main results of simulations are presented along each investigated dimension. Graphs of the complete simulation results are in the supplementary material.

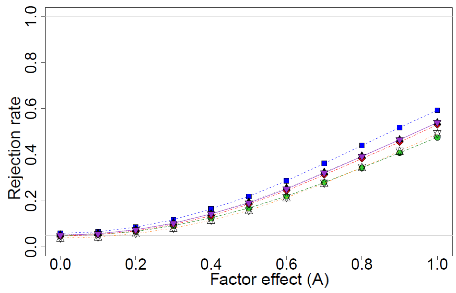
Factor effect

Power of the tests varies according to the distribution used to generate data and according to homoscedastic or heteroscedastic distribution of errors. One test could perform better in certain situations while another one could perform better in others.

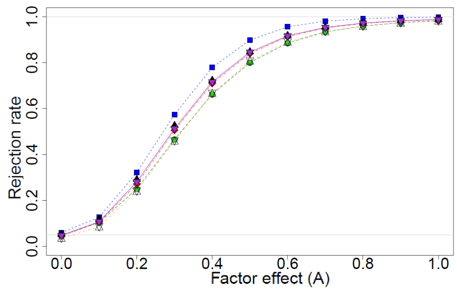
In the main factor analysis, tests show in general a monotonic increase of power when the factor effect increases (Figure 5.1). Graphs plotted along factor A and factor B are almost identical as per the way the simulation has been conducted. The graphs show the mean power along each factor. For instance, in the graph along factor A, each point represents the mean of the power observed at that specific level of factor A and at all the levels of factor B.



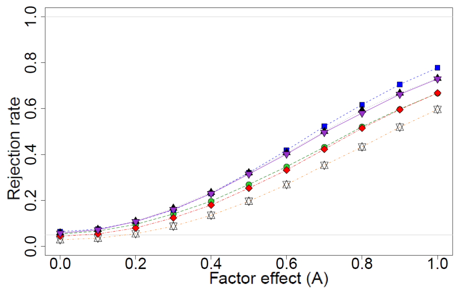
(a) Normal distribution of errors.



(b) Laplace distribution of errors.



(c) Student's t distribution of errors.



(d) Lognormal distribution of errors.

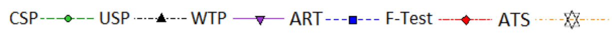


Figure 5.1: Factor A: power of test at different values of Factor Effect (x axis). Heteroscedastic case; Standard deviation = 0.5; Number of replicates = 5.

WTP, USP and F-test show often similar power in the homoscedastic case. ART test looks to perform well in case of lognormal and Laplace distributions. ATS is performing worst in case of lognormal distribution, showing a conservative behavior. Its power at factor effect=0 is below the nominal α level, thus having an impact on the type I error. Furthermore, heteroscedasticity has a clear impact on F-test's power. CSP and ATS tests have in general lower power for Laplace and lognormal distribution. The most challenging distribution for all the tests is the student's t distribution with 3 degrees of freedom (Figure 5.2). The power is quite low for all the tests, with ATS test being slightly conservative. ART test is the best performing. (See also supplementary material Figure A1, A2, A7, A8, A13, A14)

The power of the tests has been investigated along factor effect at three different levels of σ. An important aspect to consider is the ratio between factor effect and σ. The higher the ratio, the higher will be the power of the test. (See also supplementary material Figure A1, A2, A7, A8, A13, A14)

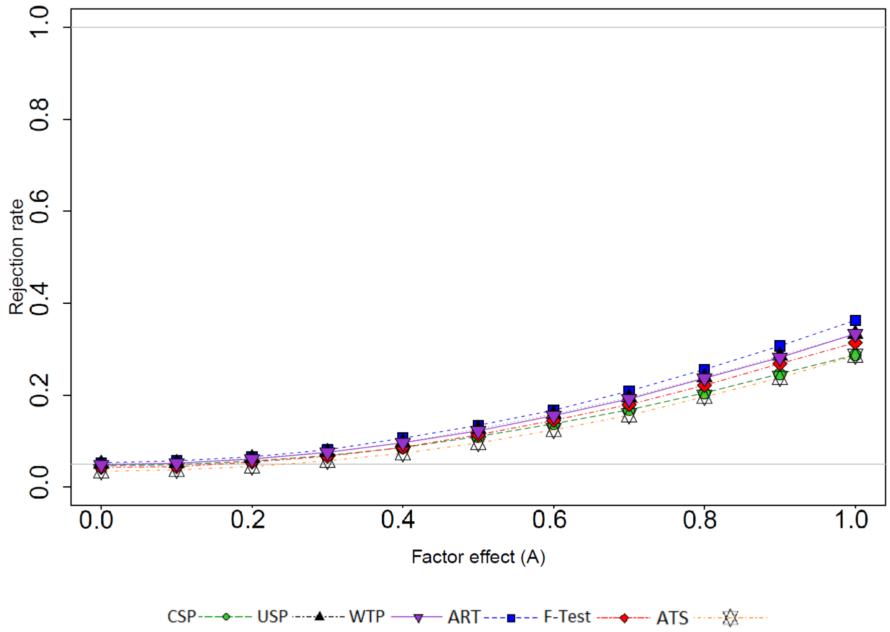


Figure 5.2: Factor A: power of test at different values of product of Factor Effect (x axis). Student's t distribution of errors with 3 d.o.f. Homoscedastic case; Number of replicates = 5.

In the interaction analysis, tests show quite different power. The power is plotted against the interaction effect. Recall that interaction is given by the product of factor effect. This fact explains the higher density of points in the left part of the graphs relative to the right part (Figure 5.3). Each point represents the mean of values observed for the same interaction effect.

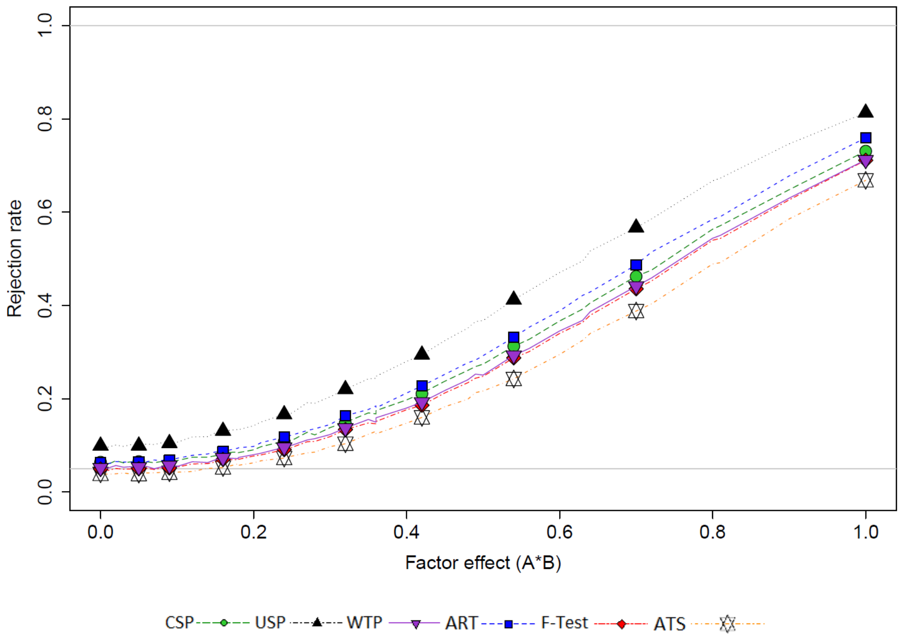
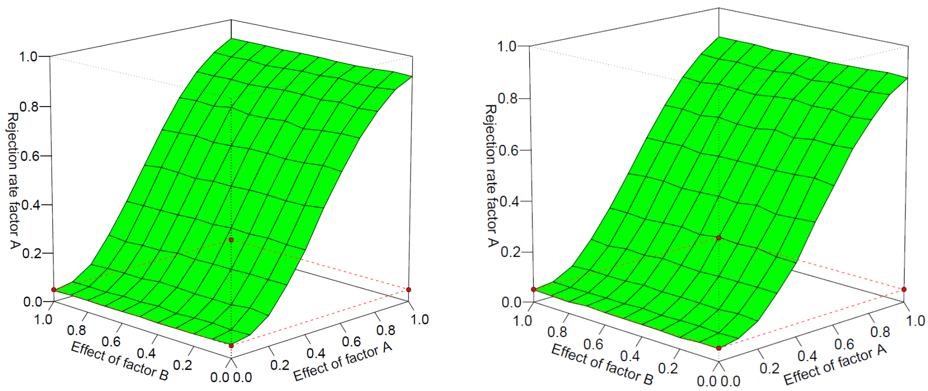


Figure 5.3: Interaction AB: power of test at different values of product of Factor Effect (x axis). Lognormal distribution of errors. Heteroscedastic case; Standard deviation = 0.25; Number of replicates = 5.

USP test shows the highest power in all the investigated cases. The problem with USP test is that its power is far above the nominal α level at interaction effect=0 (Figure 5.3), revealing a liberal behavior. There is, thus, a tradeoff between the power and the type I error in the choice of the test. All the tests show a monotonic increase of power when the interaction effect increases. In general, ATS test is performing worse and ART is performing better compared to other methods in detecting the interaction. The power of WTP, CSP and F-test is similar for the interaction analysis, even though in some cases certain tests perform better than others. (See also supplementary material Figure A5, A6, A11, A12, A17, A18)

Now that all the possible combinations of factor effects have been considered, next we investigate any possible influence on the power of the tests for the main effect of interest due to the level of the other factor effect. To that end, the power has been plotted in 3D graphs (e.g. Figure 5.4) where the surface represents the power curves of tests along factor A for each level of factor B. The red dotted line represents the α level=0.05.

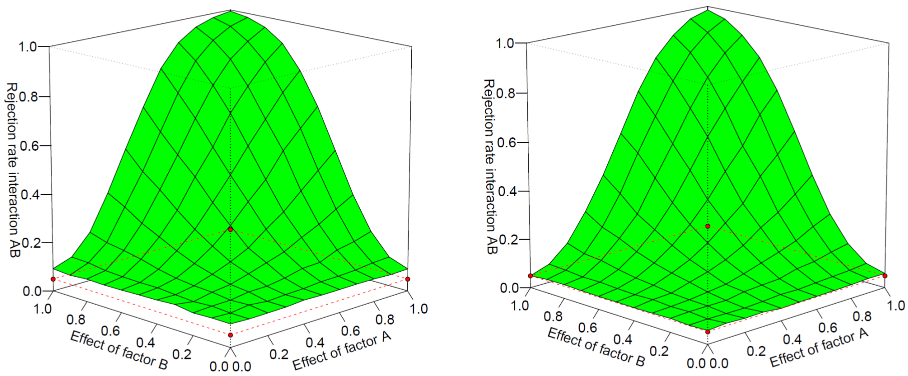


(a) WTP.   (b) CSP.

Figure 5.4: Factor A: power of WTP and ATS tests at different values of Factor Effect at each level of factor B. Lognormal distribution of errors; Homoscedastic case; Standard deviation = 0.5; Number of replicates = 5.

Power of all the investigated tests doesn't look to be affected by the level of the not investigated factor effect. In fact, the surface of power of test on main factor remains the same when the level of other factor effect increases. An example is given in Figure 5.4 with the power surface of WTP test and CSP test along factor A that keep the same level along factor B. (See also supplementary material Figure A19-A26, A43-A50, A67-A74)

The same investigation has been performed for the interaction analysis (Figure 5.5). The surface increases along the product of the factors and reaches the highest value when effect of factor A and effect of factor B are at their maximum for all the tests. Figure 5.5 confirms the liberal behavior of USP and the good performance of ART test for interaction.



(a) USP. (b) ART.

Figure 5.5: Interaction AB: power of test at different values of Factor Effect. Normal distribution of errors; Homoscedastic case; Standard deviation = 0.25; Number of replicates = 5.

Standard deviation

The variance of data has an impact on the power of the test. In the main factor analysis, power shows a monotonic decrease when the standard deviation of errors increases. In the case of normal distribution, WTP, ART and UPS tests show power comparable to the F-test, that is the highest achieved power in both homoscedastic and heteroscedastic cases. CSP appears to perform worse than others when σ increases. Non parametric tests have, in general, superior performance compared to F-test and ATS test for heteroscedastic lognormal distribution (Figure 5.6). (See also supplementary material Figure B1, B2)

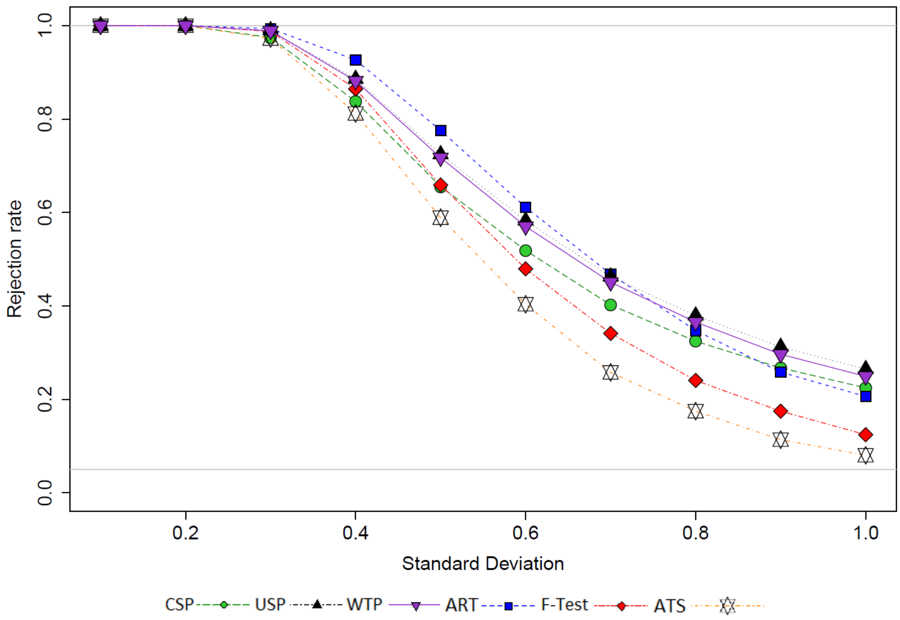


Figure 5.6: Factor A: power of test at different values of Standard Deviation (x axis). Lognormal distribution of errors. Heteroscedastic case; Factor effect = 1; Number of replicates = 5.

In the interaction analysis USP test shows the highest power in all cases (e.g. Figure 5.7). All the tests show monotonic decrease in power when σ increases. Power of WTP, CPS and ART is very similar in all cases. Heteroscedasticity reduces the power of all the tests. (See also supplementary material Figure B5, B6)

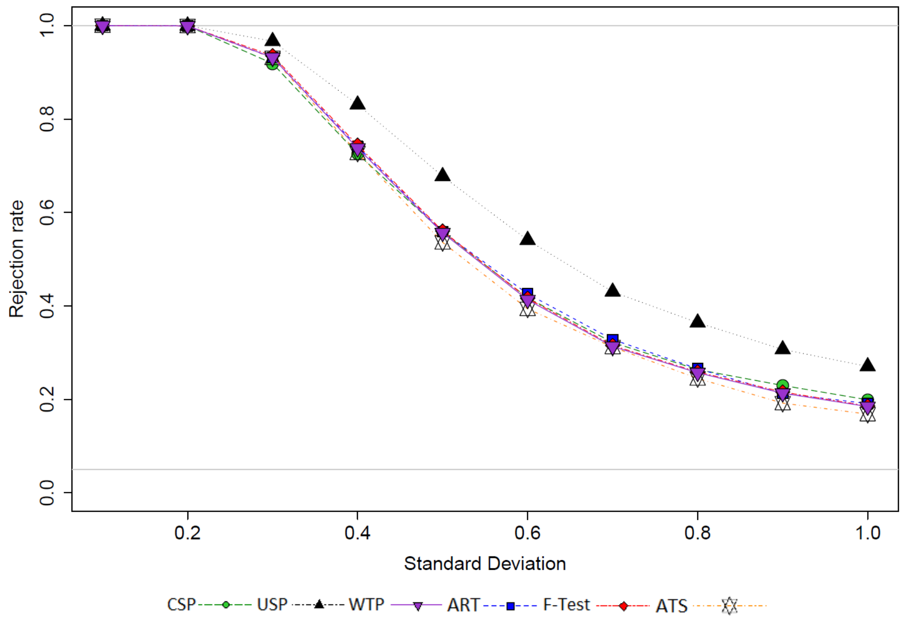


Figure 5.7: Interaction AB: power of test at different values of Standard Deviation (x axis). Normal distribution of errors. Homoscedastic case; Factor effect = 1; Number of replicates = 5.

Number of replicates

As one would expect, the number of replicates has a positive impact on the power of the tests. There is a monotonic increase of power in all cases when number of replicates increases.

In the main factor analysis, the powers of the tests are quite similar in the normal and Laplace distributions (Figure 5.8), with ART performing best in the latter case. F-test and ATS are the worst performing in the lognormal heteroscedastic case (Figure 5.9). There is a clear difference between their performances and the other tests. ART is again showing the steepest increase in power. In the case of small standard deviation σ=0.5, the power of the tests is in general high for more than 5 replicates except for student's t distribution. (See also supplementary material Figure C1, C2, C7, C8)

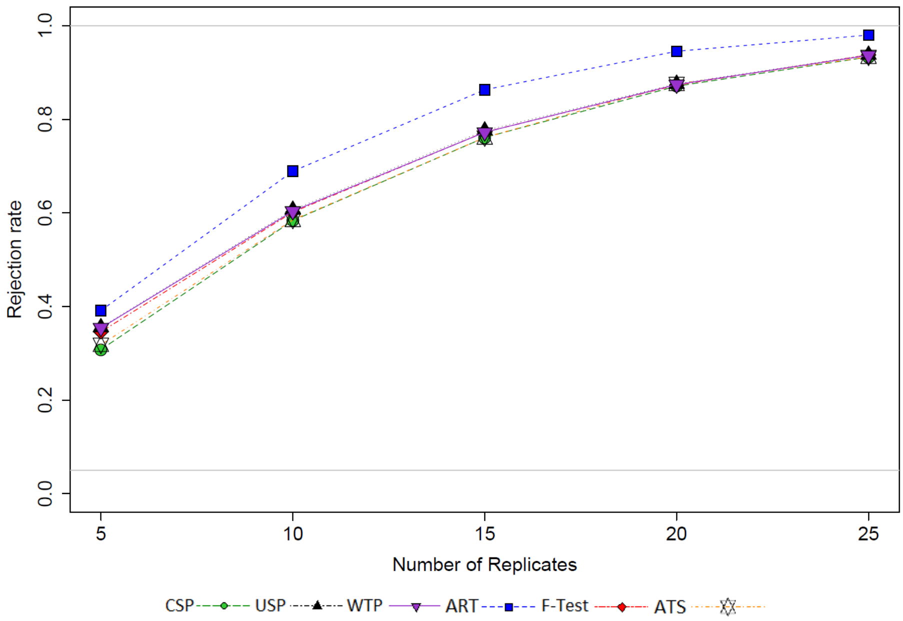


Figure 5.8: Factor A: power of test at different values of Number of Replicates (x axis). Laplace distribution of errors. Homoscedastic case; Factor effect = 1; Standard deviation = 1.

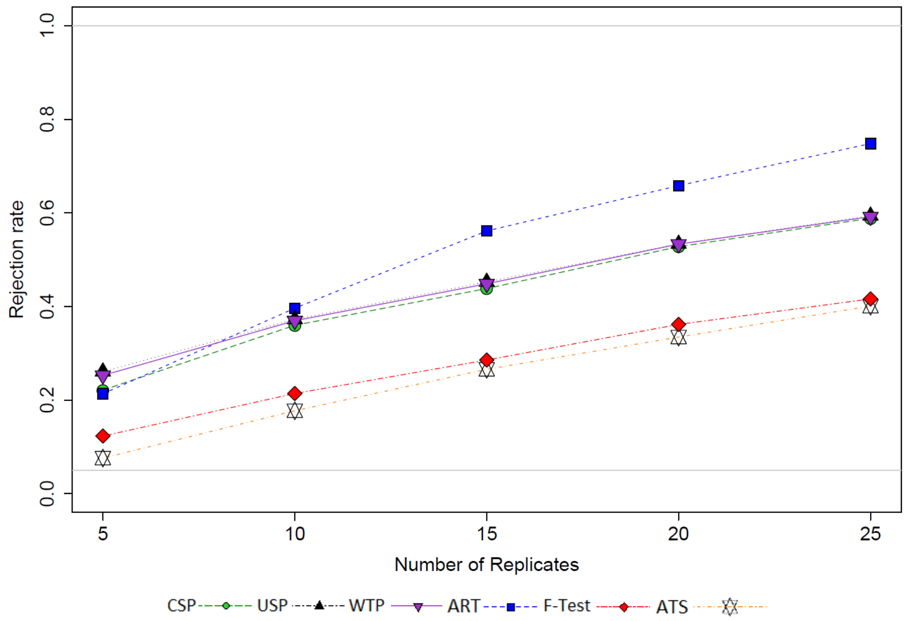


Figure 5.9: Factor A: power of test at different values of Number of Replicates (x axis). Lognormal distribution of errors. Heteroscedastic case; Factor effect = 1; Standard deviation = 1.

In the interaction analysis, ART and USP have, in general, the highest power whereas ATS often has the lowest power (Figure 5.10). (See supplementary material Figure C5, C6, C11, C12)

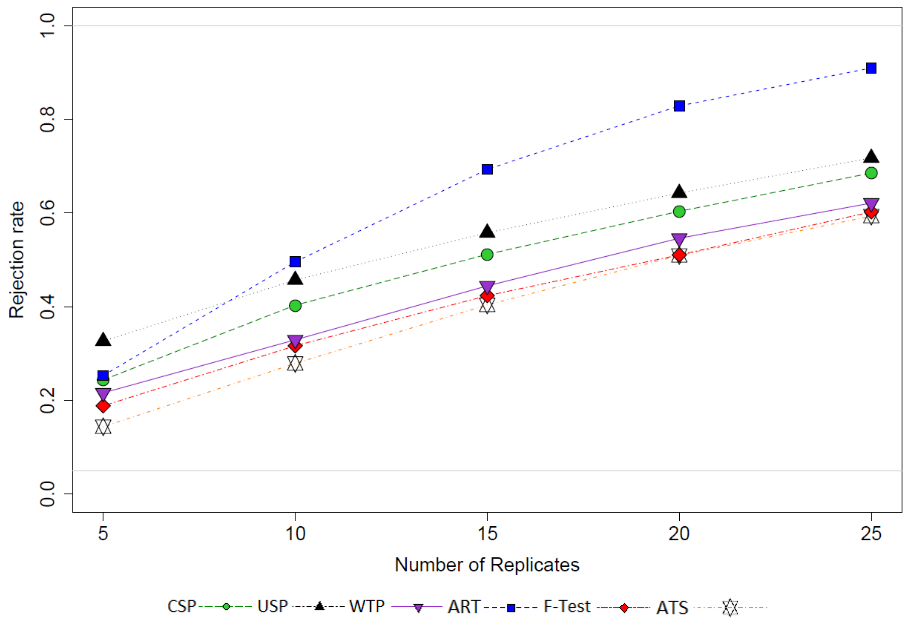


Figure 5.10: Interaction AB: power of test at different values of Number of Replicates (x axis). Lognormal distribution of errors. Heteroscedastic case; Factor effect = 1; Standard deviation = 0.5.

5.7.2 Unbalanced two-by-two designs

A simulation study to assess the power of the various methods in case of unbalanced design is performed both for homoscedastic and heteroscedastic cases. Data are generated according to the same cell mean model of section 5.7 with the only difference that number of replicates n_{ij} is not anymore constant among the factor level combinations. Because of the abundance of possible combinations of number of replicates and variances in a two-by-two factorial design, we focus the investigation to four particular settings which represent a reasonable spectrum of parameter values also used in other works [61] (Table 5.2). Settings 1 and 2 are unbalanced and homoscedastic. In setting 3 increasing sample size is combined with increasing variances (positive pairing), whereas in setting 4 increasing sample size is combined with decreasing variances (negative pairing). Two distributions are used to generate the error term ϵ, namely Laplace (symmetric) and lognormal (skewed), and the power will be investigated along effect of factors (α_i and β_j).

While ATS, WTP and ART tests can be performed without particular constraints with same packages mentioned in section 5.7, for the CSP and

Table 5.2: Simulations design for unbalanced case. In each simulation (row) factor effect is varied, while $\sigma = (\sigma_{11}, \sigma_{12}, \sigma_{21}, \sigma_{22})$ and number of replicates $n = (n_{11}, n_{12}, n_{21}, n_{22})$ are fixed.

	Investigated dimensions		
Simulation	Factor effect	σ	Number of replicates
Setting 1	0 to 1 by 0.1	(0.5, 0.5, 0.5, 0.5)	(5, 10, 20, 30)
Setting 2	0 to 1 by 0.1	(1, 1, 1, 1)	(5, 10, 20, 30)
Setting 3	0 to 1 by 0.1	(0.5, 0.5, 1, 1)	(5, 5, 10, 10)
Setting 4	0 to 1 by 0.1	(1, 1, 0.5, 0.5)	(5, 5, 10, 10)

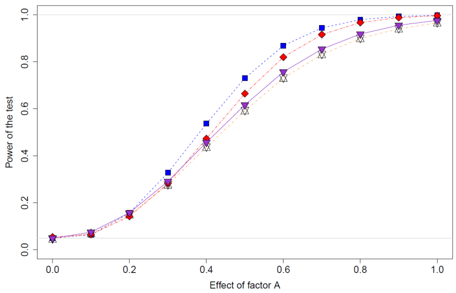
USP tests we use to the so called *fixed weight* constrained approach provided by Hahn and Salmaso [26].

Synchronized permutations have some constraints and drawbacks in the case of unbalanced design. Indeed, in each permutation the same number of units are exchanged within each pair of the considered groups. This fact implies that the combination of levels of factors with the smallest number of observations limits the number of exchanges in the other combinations. Furthermore, as the units in the same position are exchanged, some observations in the combinations with the larger sample size are never exchanged. In order to extend the use of CSP and USP tests to unbalanced designs, a few approaches have been proposed to address these issues by assigning different weights for each factor level combination when computing the test statistic. A previous study [26] shows that the *original weight* approach proposed by Basso et al. [3] does not control the nominal α level under the null hypothesis. The same happens for the *restricted weight* approach a part from the case in which $n_{11} = n_{12}$ and $n_{21} = n_{22}$. The *fixed weight* approach has been specifically developed for the latter case in which $n_{11} = n_{12}$ and $n_{21} = n_{22}$ and when testing for main effect A [26]. Under the null hypothesis and the aforementioned restrictions, this approach leads to a good adherence to the nominal α level, which makes it suitable to be used for settings 3 and 4. This approach cannot be used in setting 1 and 2, therefore a comparison of CSP and USP tests with the other methods cannot be done.

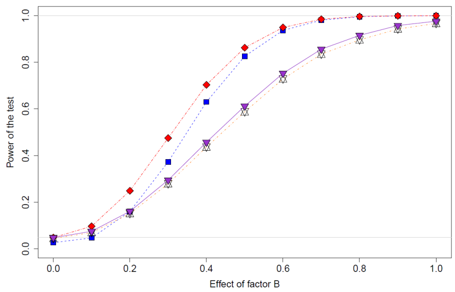
In the main factor analysis in settings 1 and 2, ART and F-test perform better than WTP and ATS, with the latter two having similar powers (Figure 5.11). Considering factor A, the two settings have an increasing number of replicates in factor levels. That is, factor A level 1 has 5 and 10 replicates, whereas level 2 has 20 and 30 replicates. If we consider factor B, the number of replicates is "mixed". That is, level 1 has 5 and 20 replicates, and level 2 has 10 and 30 replicates. We observe that ART and F-test appear to be more sensitive to the structure of the design. Indeed, their capability to detect the factor effect is greater for factor B than for factor A, whereas ATS and WTP tests have almost the same power (Figure 5.11). Concerning ART and F-test, the different number of replicates in the factor level combinations has an influence on the power of the tests for a main effect due to the level of the other factor effect. In Figure 5.12 the surface representing the power curves of ART tests as the effect of factor B varies at each level of factor A bends down, showing a different power of the test to detect the same effect at different levels of the not investigated factor. On the contrary, the curves of the power of the test on factor A are parallel. In the interaction analysis in setting 1 and 2, ART test is performing much better than the other methods. WTP, ATS and F-test have similar powers.

In the main factor analysis in setting 3, ATS and WTP tests control the nominal α level under the null hypothesis, but the others donot. Specifically, USP and F-test show very conservative behaviors, but CSP and ART tests are only slightly conservative. The ATS and WTP tests have consistently the highest power, whereas ART test performs well with the Laplace distribution of errors(Figure 5.13), but not with the lognormal distribution. In the interaction analysis, USP and F-test show very conservative behaviors. ATS and WTP tests have the highest powers which are comparable to the ART test, with the latter being slightly conservative (Figure 5.13).

In setting 4, USP, CSP, F-test and ART tests donot control the nominal α level neither in the main factor analysis nor in the interaction analysis. More specifically, they are very liberal in both cases. The four tests show the highest power but the problem is the trade off between type I error and power. WTP and ATS tests show lower power but they reveal a steady control of the nominal α level, with WTP performing better both in the main factor and in the interaction analysis (Figure 5.14).



(a) Factor A.



(b) Factor B.

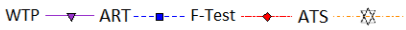
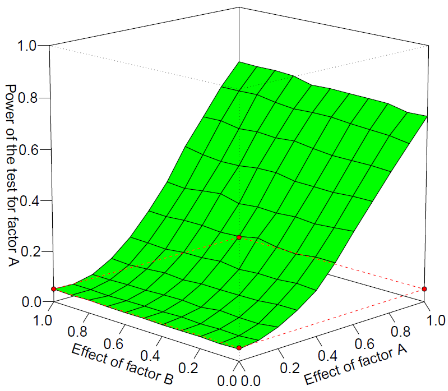
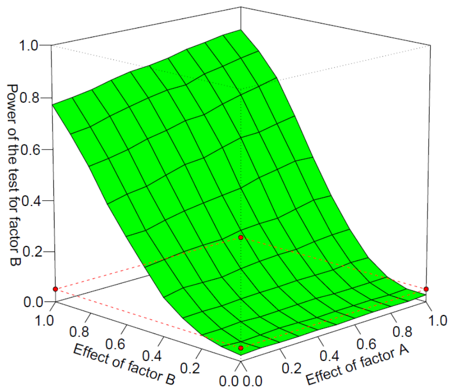


Figure 5.11: Setting 1 unbalanced case: power of test at different values of Factor Effect (x axis). Laplace distribution of errors.

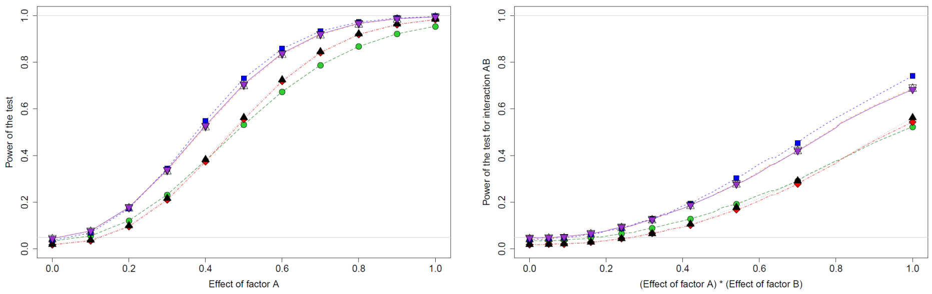


(a) ART: Factor A.



(b) ART: Factor B.

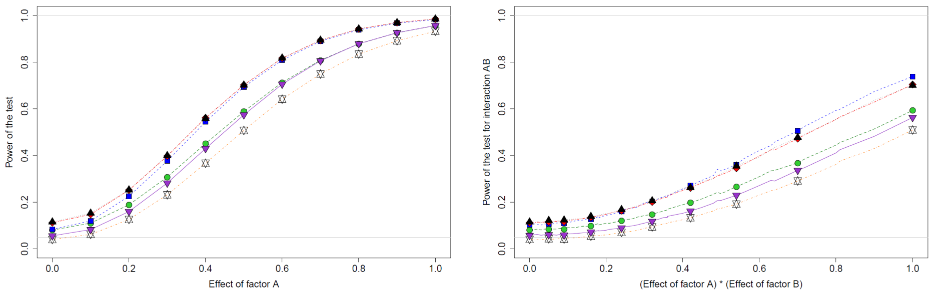
Figure 5.12: Setting 2 unbalanced case: power of ART test at different values of Factor Effect at each level of non investigated factor. Laplace distribution of errors.



(a) Factor A. (b) Interaction.

CSP USP WTP ART F-Test ATS

Figure 5.13: Setting 3 unbalanced case: power of test at different values of Factor Effect (x axis). Laplace distribution of errors.



(a) Factor A. (b) Interaction.

CSP USP WTP ART F-Test ATS

Figure 5.14: Setting 4 unbalanced case: power of test at different values of Factor Effect (x axis). Laplace distribution of errors.

5.8 Application to an Industrial Experiment

A real data set from an industrial experiment in engineering management is used to illustrate the application of the various analysis methods taking into account the power of the test suggested by the objective of the experiment.

The production system of plastic thermoformed packaging is complex and it is controlled by several factors [68]. In order to innovate the system, the impact of two factors and of their interaction has to be assessed for values of levels outside their usual range. One factor is the temperature of the process (factor A) and the other factor is the pace of production line (factor B). The response variable is the pressure needed to break the packaging by a burst test measure in bars. Historical data do not violate assumption of normality, but they reveal heteroscedasticity based on the different values of control factors. Taking into account low and high levels in the usual range of the two factors to estimate standard deviations, the scheme of heteroscedasticity is σ_{A1B1}=1.5, σ_{A1B2}=5.5, σ_{A2B1}=5.5, σ_{A2B2}=1.5. Setting the desired level of α=0.5 and the size of effect to be detected equal to 0.03 bar, a simulation has been performed to understand how many replicates should be necessary to achieve a power of the test $\mathrm{P_w} \geq 0.7$ under the assumption of a linear additive model as in equation (5.16) explaining the response. ART test allows to achieve $\mathrm{P_w} \geq 0.7$ for the main factor test with 9 replicates (Table 5.3). Impact of number of replications on interaction is less evident given the remaining parameters. The power is greater than 0.7 with all the tests at 7 replicates and beyond (Table5.4). Number of replications has been fixed at 10 and the experiment was performed. Analysis of data collected confirm the same heteroscedasticity of historical data, but they violate the assumption of normal distribution. Therefore, actual power of the tests is different from the expected power calculated by the simulation. In Table 5.5, p-values of the six statistical tests performed on the observed data are displayed. All the tests are rejecting the null hypothesis for the main factors and for interaction at level of α=0.5 with the ART test giving the smallest p-values.

Table 5.3: Main factor: power of the test. Factor effect = 0.03, normal distribution of errors, scheme of heteroscedasticity σ_{A1B1}=1.5, σ_{A1B2}=5.5, σ_{A2B1}=5.5, σ_{A2B2}=1.5.

	Tests					
Number of replicates	F test	CSP	USP	WTP	ATS	ART
5	0.35	0.28	0.35	0.35	0.31	0.48
6	0.43	0.39	0.44	0.43	0.36	0.56
7	0.49	0.45	0.49	0.49	0.44	0.62
8	0.53	0.49	0.53	0.53	0.52	0.68
9	0.60	0.57	0.60	0.59	0.55	0.72
10	0.64	0.61	0.64	0.64	0.60	0.76

5.9 Conclusions

The simulation study allowed to assess the power of some selected nonparametric methods by analyzing the same data set. Data have been generated

Table 5.4: Interaction: power of the test. Factor effect = 0.03, normal distribution of errors, scheme of heteroscedasticity σ_{A1B1}=1.5, σ_{A1B2}=5.5, σ_{A2B1}=5.5, σ_{A2B2}=1.5.

	Tests					
Number of replicates	F test	CSP	USP	WTP	ATS	ART
5	0.65	0.69	0.76	0.65	0.58	0.63
6	0.74	0.79	0.83	0.74	0.69	0.71
7	0.81	0.86	0.88	0.81	0.77	0.77
8	0.86	0.90	0.91	0.86	0.84	0.82
9	0.91	0.94	0.95	0.91	0.88	0.86
10	0.92	0.95	0.95	0.92	0.91	0.89

Table 5.5: P-values of the tests on the experimental data.

	Tests					
Factor	F test	CSP	USP	WTP	ATS	ART
Factor A	0.003	0.001	0.001	0.001	0.004	$<$ 0.001
Factor B	$<$ 0.001	$<$ 0.001	0.001	$<$ 0.001	$<$ 0.001	$<$ 0.001
Interaction AB	0.009	0.003	0.001	0.006	0.02	$<$ 0.001

using a linear additive model for a two-way two-levels design with interaction given by the product of factor level effects. Such model is common for practitioners in industrial experimentation.

The study reveals that certain methods of analysis perform better than others depending on the dataset and on the objective of the analysis. As a consequence, there does not emerge a unique approach in the design phase of the experiment, but various aspects have to be taken into account simultaneously as shown in the example provided in Section 5.8. The three dimensions (factor effect, standard deviation, number of replicates) along which the investigation has been conducted have impacts on the power of the tests. Furthermore, the study allowed to bring out some interesting information.

Concerning the balanced case, ART test is an overall well performing test both for the main factor and the interaction analysis. It has superior power in almost all the settings, and it maintains the α level well. WTP test is a good test as well both for main factor and interaction analysis. These findings corroborate with previous studies ([25, 61]). ATS test's performance is at an average level for main factor analysis, showing slightly lower power in the heteroscedastic case and resulting conservative in some cases. In the interaction analysis ATS is performing worse relative to the other tests. In the main factor analysis, USP is performed well in most of the situations and its power is similar to that of WTP. In the interaction analysis, USP is performed best compared to the other tests in all the cases, but it doesnot maintain the nominal α level. The liberal behavior is a limitation of this test because of the tradeoff between power and type I error. CSP doesnot reveal any liberal behavior in the interaction analysis and its power is similar to that of WTP. In agreement to results of Hahn et al. [25], in the main factor analysis CSP has, in general, slightly lower power for Laplace and lognormal distribution relative to the other tests. In the main factor analysis, there is no influence on the power of the tests due to the other factor's level. The results for CSP and USP are also in agreement with results of a previous study ([3]).

In the unbalanced case, WTP and ATS tests are the only ones that appear to be reliable tests in all the scenarios considered. Indeed, they control the α level and maintain the power level when the number of replicates is switched between the factor level combinations. CSP and USP tests can be used due to the *fixed weight* approach only when $n_{11} = n_{12}$ and $n_{21} = n_{22}$, but they donot control the α level, resulting in conservative decisions in the so-called *positive pairing* heteroscedastic setting and liberal decision in the *negative pairing* for the main factor analysis as well as the interaction analysis, unlike the ART test. In the homoscedastic case, ART test performs better than WTP and ATS, and its power varies when the number of replicates is switched between the factor level combination. In the main factor analysis, ART test reveals an influence on its power due to the level of the other factor in some configurations.

The methods of analysis compared in this study were seen to be effective and reliable depending on the situation at hand. Nevertheless there are some issues that need to be further investigated or that could pose practical problems. Warning messages have been obtained from some of the packages for those cases in which all the observations in a factor level combination have the same value. In such cases, WTP and ATS tests cannot be done because of the way the test statistics are computed, whereas CSP, USP and ART tests do not show such limitations. Furthermore, this study is limited to the simulation designs described in sections 5.7. Investigating different designs could generate more information and guidance to practitioners. For example, the case of repeated measures or larger designs like three-by-three crossed factorial design are frequently encountered designs. Specifically, while CSP and USP can be implemented in designs with a higher number of levels in the balanced case [4], it is difficult to deal with synchronized permutation methods in unbalanced three-by-three design and it could be direction for further research.

Chapter 6

Multivariate Nonparametric tests: Two-way Designs for Industrial Experiments

6.1 Introduction

Rationale of practitioners in industrial experimentation is often based on factorial designs. A response variable that measures the phenomenon under investigation is observed. Practitioners assess the impact on response variable of factors they can control during the experiments according to different levels. Design of Experiment (DOE) is used in current research in different fields of engineering, as for instance for machines' production process [44], for industrial production practices [38, 79], for bio-fuel production [83], for alternative raw materials experimentation [43], or for the use of Coordinate Measuring Machines for quality control [72]. The two-way crossed factorial design is a common design used in the exploratory phase in industrial experiments. This design allows to investigate the impact on response variable of each factor and of interaction between factors, thus allowing to assess whether factors or interaction are significant or not according to the assumed model.

In many industrial applications (and applied research fields) it is common the need to compare multivariate population obtained in advanced factorial designs. There are manufacturing processes where treatments or control factors in production processes impact on several relevant variables simultaneously [68, 51, 15]. In these cases an overall test is useful to determine for instance whether there is a significant difference on final product or not. Observed data are usually analized using the multivariate analysis of variance

(MANOVA) methods. Unfortunately parametric methods rely on assumptions such as multivariate normality and covariance homogeneity, but these prerequisites may be not realistic for several real problems.

How to overcome the violation of MANOVA assumptions has been investigated and nonparametric methods for multivariate inferential tests have been developed. One of the approaches is based on the generalization to the multivariate case [58, 12] of the univariate comparison between the group-wise distribution functions F_i and the reference distribution function $H = \frac{1}{N}\sum_i n_i F_i$ that is the pooled distribution [45, 46]. In these cases the null hypotheses are formulated in terms of distribution functions. Other approaches are rank-based as for instance in [28, 5, 6, 30, 29] but they are for large number of factor levels or for large sample size.

I propose a novel nonparametric approach based on NonParametric Combination (NPC) [62] applied to Synchronized Permutation (SP) tests [3] for two-way crossed factorial design assuming a linear additive model. Indeed, the linear additive model interpretation well adapts to the industrial production environment because of the way control of production machineries is implemented. This approach overcomes the shortcomings of MANOVA with the only mild condition of the data set to be analyzed taking values on a multi-dimensional distribution belonging to a nonparametric family of non-degenerate probability distributions. It well works with even only two levels per factor and a small sample size. The case of small sample size reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. Furthermore it allows to formulate test hypotheses in more familiar terms for practitioners such as factor effect size. Indeed, I agree with Lakens [48] that "Effect sizes are the most important outcome of empirical studies. Most articles on effect sizes highlight their importance to communicate the practical significance of results".

A simulation design with fixed factor effects δ and fixed variance σ of data set distributions have been performed in order to evaluate the rejection rate of the NPC applied to SP under alternative Hypothesis H_1 in the range of interest of significance levels $0 \leq \alpha \leq 0.1$, and in order to compare it with the classical MANOVA test.

A real case study is useful to highlight the benefits of the adoption of the herein presented nonparametric approach in industrial experiments with a small sample size and non-normal data distribution. A two-way two-level design is used to understand whether two control factors and their interaction were significant or not in a project of innovation of the production system of a thermoformed packaging.

Results presented in this chapter have been published [2].

6.2 Model, Hypotheses and Statistics of Synchronized Permutations

The linear additive model is a common model in the industrial environment. It reflects the logic adopted by practitioners in many cases, it's easy to understand and well adapts to the most common control models of machineries. I assume the linear additive model of a balanced two-way factorial crossed design:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, n, \tag{6.1}$$

where μ is the overall mean, α_i is the effect of level i of factor A, β_j is the effect of level j of factor B, $(\alpha\beta)_{ij}$ is the effect of interaction between factor A at level i and factor B at level j, I and J are the number of levels of factor A and B respectively, and ϵ_{ijk} is the error term. The number of replicates of the balanced design is n and the mean of error term is $E(\epsilon_{ijk}) = 0$ for each factor level combination. The total number of observations is $N = \sum_i \sum_j n = I \cdot J \cdot n$

The side conditions are:

$$\sum_i \alpha_i = 0; \quad \sum_j \beta_j = 0; \quad \sum_i (\alpha\beta)_{ij} = 0 \ \forall j; \quad \sum_j (\alpha\beta)_{ij} = 0 \ \forall i. \tag{6.2}$$

The null hypotheses of no-main effect of factor A, no-main effect of factor B and no-interaction effect between factors A and B are:

$$\begin{aligned} H_0^{(A)} &: \quad \alpha_i = 0 \quad \forall i, \\ H_0^{(B)} &: \quad \beta_j = 0 \quad \forall j \quad \text{and} \\ H_0^{(AB)} &: \quad (\alpha\beta)_{ij} = 0 \quad \forall i, j, \end{aligned} \tag{6.3}$$

respectively. In vector notation, the data can be written as $\boldsymbol{Y} = (Y_{ijk})' = (Y_{111}, \dots, Y_{IJn})'$. And the null hypotheses can be written in terms of contrasts as:

$$\begin{aligned} H_0^{(A)} &: \quad \boldsymbol{C}_A \boldsymbol{\mu} = \boldsymbol{0}, \\ H_0^{(B)} &: \quad \boldsymbol{C}_B \boldsymbol{\mu} = \boldsymbol{0} \quad \text{and} \\ H_0^{(AB)} &: \quad \boldsymbol{C}_{AB} \boldsymbol{\mu} = \boldsymbol{0}, \end{aligned} \tag{6.4}$$

where $\boldsymbol{\mu} = (\mu_{11}, \dots, \mu_{IJ})'$, $\mu_{ij} = E(Y_{ijk}) = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ and $\boldsymbol{C}_M$ is a contrast matrix for $M \in \{A, B, AB\}$

The synchronized permutation methods is herein illustrated. Synchronized permutation is developed along two basic concepts. The first is that permutations of observations between two levels of a factor can be made only holding the level of remaining factors in the model constant. For instance, consider the case of a two-way design with factor level combinations A_1B_1, A_1B_2, A_2B_1 and A_2B_2. To test the significance of main effect A, observations will be exchanged between groups A_1B_1 and A_2B_1, and between A_1B_2 and A_2B_2. That is, the level of B is kept constant when performing the test on factor A, in the former case level is 1, in the latter it is 2. The second basic concept in synchronized permutation is exchanging the same number of units within each pair of the considered groups [3].

According to Basso et al. [3] the test statistics for the main factor A in the two-way design is:

$$\begin{aligned} T_A &= \sum_{i<s}\left[\sum_j T_{is|j}\right]^2, \quad \text{where} \\ T_{is|j} &= \sum_k Y_{ijk} - \sum_k Y_{sjk}, \quad i,s \in \{1,\ldots,I\}; \quad j \in \{1,\ldots,J\} \end{aligned} \tag{6.5}$$

The outer sum is made over all possible pairs of levels $1 \leq i \leq s \leq I$ and the inner sum is squared to avoid the cancellation of any of the contributions of effects of factor A.

Similarly for factor B:

$$\begin{aligned} T_B &= \sum_{j<h}\left[\sum_i T_{jh|i}\right]^2, \quad \text{where} \\ T_{jh|i} &= \sum_k Y_{ijk} - \sum_k Y_{ihk}, \quad i \in \{1,\ldots,I\}; \quad j,h \in \{1,\ldots,J\} \end{aligned} \tag{6.6}$$

The statistics for interaction between factor A and factor B is given by the summation of two contributions along the two factors:

$$\begin{aligned} T_{AB} &=^a T_{AB} +^b T_{AB}, \quad \text{where} \\ {}^aT_{AB} &= \sum_{i<s}\sum_{j<h}\left[T_{is|j} - T_{is|h}\right]^2, \quad \text{and} \\ {}^bT_{AB} &= \sum_{j<h}\sum_{i<s}\left[T_{jh|i} - T_{jh|s}\right]^2, \quad i,s \in \{1,\ldots,I\}; \quad j,h \in \{1,\ldots,J\} \end{aligned} \tag{6.7}$$

The statistics for main factors and interaction are uncorrelated.

Then p-value is calculated as the proportion of permutations for which test statistics of permuted data set are greater or equal to the test statistic of the original data set.

There are two ways to obtain a synchronized permutation, namely Constrained Synchronized Permutation (CSP) [69] and Unconstrained Synchronized Permutation (USP) [69].

Constrained Synchronized Permutation (CSP)

The CSP is computationally less intensive compared to USP. In the CSP the same permutation is applied in all couples of groups given the initial order of observations. For instance, in the two way design if a permutation consists in exchanging the second observation of group A_iB_j with the first observation of group A_sB_j when testing main effect A, then same permutation has to be applied to groups A_iB_h and A_sB_h. It is recommended to randomize observations within each group at the beginning before performing the permutation test.

As a result of the application of the same permutation between all possible pairs of groups, the number of possible ways to exchange units depends only on number of replicates n in the balanced design. The total number of possible permutations of CSPs is:

$$C_{csp} = \binom{2n}{n} \tag{6.8}$$

Thus, according to the way the p-value is calculated, the minimum achievable significance error is $\alpha_{min} = 2 \times (C_{csp})^{-1}$. If n is too small, CSP could give a minimum achieved significance level higher than the desired type I error.

Unconstrained Synchronized Permutation (USP)

The USP is computationally more intensive compared to CSP. USP, unlike CSP, can apply different permutations in the various pairs of groups. However, the basic principle of synchronized permutations of exchanging the same number of observations has to be respected. The algorithm provided by Basso et al. [3] guarantees the values of the test statistic to be equally likely. This procedure allows to overcome those cases in which the test statistic is not uniformly distributed.

The total number of possible permutations of USPs depends on a larger number of parameters of dataset respect to CSP. The formula is more complex and there are two cases:

$$\begin{aligned} C^o_{USP} &= \sum_{\nu=0}^{(n-1)/2} \binom{n}{\nu}^{J\times I(I-1)} && \text{when n is odd,} \\ C^e_{USP} &= \sum_{\nu=0}^{n/2-1} \binom{n}{\nu}^{J\times I(I-1)} + \left[\frac{1}{2}\binom{n}{n/2}^{2J}\right]^{I(I-1)/2} && \text{when n is even,} \end{aligned} \tag{6.9}$$

where ν is the number of units exchanged between two groups. The cardinality of the permuted statistics rapidly increases with n, I and J. The minimum significance level that can be achieved is proportional to the inverse of the cardinality in Equation 6.9. Thus USP is not expected to suffer of a minimum significance level higher than the desired type I error. However, as USP is computationally more intensive compared to CSP, it is recommended in the case of small number of replicates.

6.3 Model, Hypotheses and Statistics of Non-Parametric Combination (NPC)

The multivariate nonparametric tests are the core of this study. In previous section and in chapter 5, I illustrate the Synchronized Permutation tests whic are a useful instrument in case of violation of assumptions of univariate parametric tests such as F-test for ANOVA. In this section I introduce the NonParametric Combination which is a natural extension of permutation testing to a variety of multivariate problems. Permutation tests are, in general, distribution-free and non-parametric [23], and have good properties such as exactness, unbiasedness and consistency [62, 34].

To illustrate the NPC we assume the same linear additive model and use the same notation as in Section 6.2. The notation has to be extended to the multivariate case introducing P observed variables that can be independent or dependent. Because of the objective of this study, we focus on continuous variables. Let us denote a P-dimensional data set by $\boldsymbol{Y} = \{\boldsymbol{Y}_{i,j,k}, \quad i = 1,\dots,I, j = 1,\dots,J, k = 1,\dots,n\} = \{Y_{i,j,k,p}, \quad i = 1,\dots,I, j = 1,\dots,J, k = 1,\dots,n, p = 1,\dots,P\}$. According to the extended notation, the multivariate linear additive model of a balanced two-way factorial crossed design is:

$$\boldsymbol{Y}_{i,j,k} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + (\boldsymbol{\alpha\beta})_{ij} + \boldsymbol{\epsilon}_{ijk}, \; i = 1,\dots,I; \; j = 1,\dots,J; \; k = 1,\dots,n, \tag{6.10}$$

where, in the multivariate case, $\boldsymbol{\mu}$ is the vector of overall means, $\boldsymbol{\alpha}_i$ is the vector of effects of level i of factor A, $\boldsymbol{\beta}_j$ is the vector of effects of level j of factor B, $(\boldsymbol{\alpha\beta})_{ij}$ is the vector of effects of interaction between factor A at level i and factor B at level j, I and J are the number of levels of factor A and B respectively, and $\boldsymbol{\epsilon}_{ijk}$ is the vector of error terms. The vector of the means of error terms is $E(\boldsymbol{\epsilon}_{ijk}) = \mathbf{0}$ for each factor level combination, and $\boldsymbol{\Sigma}$ is the variance/covariance matrix of the P observed variables.

Adapted from Pesarin and Salmaso [62] to the design of interest in this study, main assumptions regarding the data structure, hypotheses being tested in NPC contexts, and set of partial tests are:

- (**i**) The response $\boldsymbol{Y}$ takes its values on a P-dimensional distribution, $D_{i,j} \in \mathrm{D}, i = 1, \ldots, I, j = 1, \ldots, J$, belonging to a (possibly not specified) nonparametric family D of non-degenerate probability distributions.

- (**ii**) The null hypothesis refers to equality of effect vectors of the P variables in the I groups for factor A, the J groups for factor B and the $I \cdot J$ groups for interaction between factor A and factor B:

$$\begin{aligned} H_0^{(A)} &: \quad \boldsymbol{\alpha}_i = \mathbf{0} \quad \forall i, \\ H_0^{(B)} &: \quad \boldsymbol{\beta_j} = \mathbf{0} \quad \forall j \quad \text{and} \\ H_0^{(AB)} &: \quad (\boldsymbol{\alpha\beta})_{ij} = \mathbf{0} \quad \forall i, j, \end{aligned} \tag{6.11}$$

The null hypotheses $H_0^{(A)}$ and $H_0^{(B)}$ imply that the P-dimensional data vectors in $\boldsymbol{Y}$ are exchangeable with respect to the I and J groups respectively.

Considering factor A, $H_0^{(A)}$ is supposed to be properly and equivalently broken down into P sub-hypotheses $H_{0p}^{(A)}$, $p = 1, \ldots, P$, each appropriate for a partial (univariate) aspect of interest. Therefore, $H_0^{(A)}$ (multivariate) is true if all the $H_{0p}^{(A)}$ are jointly true; and so it may be written as:

$$H_0^{(A)} : \left\{ \bigcap_{p=1}^{P} H_{0p}^{(A)} \right\} \tag{6.12}$$

$H_0^{(A)}$ is also called *global* or *overall null hypothesis* for factor A, and $H_{0p}^{(A)}$, $p = 1, \ldots, P$ are called the *partial null hypotheses*. Similarly $H_0^{(B)}$ and $H_0^{(AB)}$ are supposed to be properly and equivalently broken down into P sub-hypotheses, giving:

$$\begin{aligned} H_0^{(B)} &: \left\{ \bigcap_{p=1}^{P} H_{0p}^{(B)} \right\}, \\ H_0^{(AB)} &: \left\{ \bigcap_{p=1}^{P} H_{0p}^{(AB)} \right\} \end{aligned} \tag{6.13}$$

- (**iii**) Considering factor A, the alternative hypothesis states that at least one of the partial null hypotheses $H_{0p}^{(A)}$ is not true. Hence, the alternative may be represented by the union of P partial alternatives hypotheses,

$$H_1^{(A)} : \left\{ \bigcup_{p=1}^{P} H_{1p}^{(A)} \right\}, \tag{6.14}$$

stating that $H_1^{(A)}$ is true when at least one partial alternative hypotheses $H_{1p}^{(A)}$ is true. In this context, $H_1^{(A)}$ is called the *global* or *overall alternative hypothesis*. Based on the same rationale, we have:

$$\begin{aligned} H_1^{(B)} &: \left\{ \bigcup_{p=1}^{P} H_{1p}^{(B)} \right\}, \\ H_1^{(AB)} &: \left\{ \bigcup_{p=1}^{P} H_{1p}^{(AB)} \right\} \end{aligned} \tag{6.15}$$

- **(iv)** $\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{Y})$ represents a P-dimensional vector of test statistics, $P \geq 2$, in which the p-th component $T_p = T_p(\boldsymbol{Y})$, $p = 1, \ldots, P$, represents the non-degenerate p-th *partial univariate test* appropriate for testing sub-hypothesis H_{0p} against H_{1p}. In the NPC context, without loss of generality, all partial tests are assumed to be marginally unbiased, consistent and significant for large values.

The above set of mild conditions should be jointly satisfied. Concerning the partial univariate test, we note that Synchronized Permutation test respects requirements of point (iv) (for more details see [3]). Without loss of generality, from here on we intend the partial univariate test and related statistics to be the Synchronized Permutation test and its statistics.

When developing a multivariate hypothesis testing procedure, a global answer including several response variables is required, and the main point is how to combine the information related to the P variables into one global test. The key idea in the NPC to test the global null hypoteses $H_0^{(A)}$, $H_0^{(B)}$ and $H_0^{(AB)}$ is to combine through an appropriate combining function the partial (univariate) tests which are focused on the p-th component variable. Basically, NPC approach corresponds to a method of analysis made up of two phases. In the first phase the univariate permutation tests are performed. In the second phase the p-values obtained in the first phase are combined in one second-order global (multivariate) test:

$$T'' = \phi\left(\lambda_1, \ldots, \lambda_P\right) \tag{6.16}$$

where T'' is the multivariate statistic, ϕ is the combining function and λ_p, $p = 1, \ldots, P$ is the p-value of the p-th partial univariate test. The test is performed by a continuous, non-increasing and univariate real function $\phi : (0,1)^P \to \mathcal{R}^1$.

6.3.1 Combining Functions

Combining functions are a key of the success of this multivariate nonparametric tests. Various combining functions can be suitable for this purpose, but according to Pesarin and Salmaso [62], in order to be suitable for test combination all combining functions ϕ must satisfy the following properties (see also [63, 65, 64] and [24]):

- (**i**) The function ϕ must be non-increasing in each argument: $\phi(\ldots, \lambda_p, \ldots) \geq \phi(\ldots, \lambda'_p, \ldots)$ if $\lambda_p < \lambda'_p$, $p \in \{1, \ldots, P\}$.
- (**ii**) The function ϕ must attain its supremum value $\bar{\phi}$, possibly not finite, even when only one argument attains zero: $\phi(\ldots, \lambda_p, \ldots) \to \bar{\phi}$ if $\lambda_p \to 0$, $p \in 1, \ldots, P$.
- (**iii**) $\forall \alpha > 0$, the critical value T''_α of every ϕ is assumed to be finite and strictly smaller than $\bar{\phi} : T''_\alpha < \bar{\phi}$.

In the simulation study we present in Sections 6.5 and 6.5.1, we selected and compared performances of three combining functions that satisfy the required properties, namely:

- (**i**) The Fisher *omnibus* combining function is based on the statistic:

  $$T''_F = -2 \cdot \sum\nolimits_p \log(\lambda_p) \tag{6.17}$$

  If the P partial test statistics are independent and continuous, then in the null hypothesis T''_F follows a central χ^2 distribution with $2P$ degrees of freedom.

- (**ii**) The Liptak combining function is based on the statistic:

  $$T''_L = \sum\nolimits_p \Phi^{-1}(1 - \lambda_p) \tag{6.18}$$

  where Φ is the standard normal cumulative distribution function (CDF). If the P partial tests are independent and continuous, then in the null hypothesis T''_L is normally distributed with mean 0 and variance P (see [50]).

- **(iii)** The Tippett combining function is based on the statistic:

$$T''_T = \max_{1 \leq p \leq P}(1 - \lambda_p) \tag{6.19}$$

significant for large values. Its null distribution, if the P tests are independent and continuous, behaves according to the largest of P random values from the uniform distribution in the open interval $(0, 1)$.

6.4 A Two Phase Algorithm to Combine NPC and SP

I herein illustrate the two phases algorithm to perform the NPC test in the framework of the Conditional Monte Carlo Procedure (CMCP). At this stage, once defined data set structure, null and alternative hypotheses, univariate test statistic, combining functions and various assumptions and properties required in the NPC context, the two phase algorithm allows to perform the inferential test. I resort to CMCP because in most real problems computational difficulties arise in calculating the conditional permutation space when the sample size is large enough, therefore it could be not possible to calculate the exact p-value λ_p of the observed statistic T_p^{obs} in a reasonable amount of time. It is worth noting that in the multivariate data set CMCP apply permutations of individual data vectors, so that all underlying dependence relations which are present in the component variables are preserved.

For the sake of clearness and simplicity, the algorithm is presented referring to a general test of hypothesis. The reader should take in mind that it has to be repeated three times in a two-way factorial design to test the three global null hypotheses $H_0^{(A)}$, $H_0^{(B)}$ and $H_0^{(AB)}$.

The first phase of the algorithm to perform NPC test is devoted to the estimation of P-variate distribution of $\boldsymbol{T}$ and the p-values of the univariate tests:

- (i) Calculate the P-dimensional vector of the observed values of test statistics $\boldsymbol{T} : \boldsymbol{T}^{obs} = \boldsymbol{T}(\boldsymbol{Y}) = \left[T_p^{obs} = T_p(\boldsymbol{Y}),\ p = 1, \ldots, P\right]$.

- (ii) Consider a random permutation $\boldsymbol{Y}^*$ of $\boldsymbol{Y}$ and calculate the vector of statistics $\boldsymbol{T}^* = \boldsymbol{T}(\boldsymbol{Y}^*) = \left[T_p^* = T_p(\boldsymbol{Y}^*),\ p = 1, \ldots, P\right]$.

- (iii) Repeat the previous step C times independently. The set of CMC results $\{\boldsymbol{T}_c^*, c = 1, \ldots, C\}$ is thus a random sampling from the permutation P-variate distribution of vector of test statistics $\boldsymbol{T}$.

- (iv) According to Synchronized Permutation test, the p-values of the observed values are calculated in each univariate test as the proportion of permutations for which test statistics of permuted data set are greater or equal to the test statistic of the original data set:

$$\hat{\lambda}_p^{obs} = \sum_{c=1}^{C} \mathbb{I}(T_{cp}^{*} \geq T_p^{obs})/C, \quad p = 1, \ldots, P, \tag{6.20}$$

  where $\mathbb{I}(\cdot)$ is the indicator function. The result is $\{\hat{\lambda}_1^{obs}, \ldots, \hat{\lambda}_P^{obs}\}$.

- (v) The p-values of each of the C elements of the set of permutations carried out at point (iii) are calculated in each univariate test in similar way as point (iv). The result is $\{\hat{\lambda}_{c1}^{*}, \ldots, \hat{\lambda}_{cP}^{*}\}$, $c = 1, \ldots, C$.

The second phase of the algorithm to perform NPC test is devoted to the combination of results of the first phase to compute a second-order global (multivariate) test for the overall null hypothesis:

- (i) The combined observed value of the second-order test is calculated by applying a combining function to the p-values of the observed values:

$$T''^{obs} = \phi(\hat{\lambda}_1^{obs}, \ldots, \hat{\lambda}_P^{obs}). \tag{6.21}$$

- (ii) The c-th combined value of the P-dimensional vector of p-values of the c-th element of the set of permutations is then calculated by:

$$T_c''^{*} = \phi(\hat{\lambda}_{c1}^{*}, \ldots, \hat{\lambda}_{cP}^{*}), \quad c = 1, \ldots, C. \tag{6.22}$$

- (iii) Hence, the p-value of the combined test T'' is estimated as

$$\hat{\lambda}_\phi'' = \sum_{c=1}^{C} \mathbb{I}(T_c''^{*} \geq T''^{obs})/C. \tag{6.23}$$

- (iv) If $\hat{\lambda}_\phi'' \leq \alpha$, the global null hypothesis H_0 is rejected at significance level α.

6.5 The Simulations Campaign

Simulation studies are in general the most common way to evaluate the respect of the α level of inferential tests under the null hypothesis H_0 and the rejection rate under the alternative hypothesis H_1. A Monte Carlo simulation study is performed to evaluate the performances of the application of NPC methods to the SP tests described in Sections 6.3 and 6.2 respectively.

Data set for simulation are generated according to the cell mean model of a multivariate two-way balanced crossed factorial design (factor A and factor B). Consistently with notation in Section 6.3:

$$\boldsymbol{Y}_{i,j,k} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + (\boldsymbol{\alpha\beta})_{ij} + \boldsymbol{\epsilon}_{ijk}, \tag{6.24}$$

$i = 1, \dots, I$, $j = 1, \dots, J$, $k = 1, \dots, n$, $p = 1, \dots, P$, where I and J are the number of levels of factor A and B respectively, n is the number of replicates and P is the number of response variables. As the factorial design is balanced, number of replicates refers to each factor level combination. In this setup, the vector of overall means $\boldsymbol{\mu} = \mathbf{0}$ and the interaction is given by the product of effect of the two factors.

Four distributions are used to generate the error term ϵ. Three are symmetric: normal, Laplace and student's t with 2 degrees of freedom (d.o.f.). One is skewed: lognormal. We consider only homoscedastic case.

Some parameters in the model are fixed:

- The factor effect is $\delta = 1$ for both factors. According to the adopted simulation design, the maximum difference between the means of two levels due to a single factor effect is δ. In the case of two levels of factor A we have $\alpha_1 = 0.5$ and $\alpha_2 = -0.5$, while in the case of three levels we set $\alpha_1 = 0.5$, $\alpha_2 = 0$ and $\alpha_3 = -0.5$. The same for factor B: $\beta_1 = 0.5$, $\beta_2 = -0.5$, and $\beta_1 = 0.5$, $\beta_2 = 0$, $\beta_3 = -0.5$ in the case of two and three levels respectively.

- The variance of distributions is fixed as well: $\sigma^2 = 1$

Some other parameters in the model are varied:

- The number of levels of factors: $I, J = 2, 3$. We consider two possible settings: $(I, J) \in \{(2, 2), (3, 3)\}$, so the two factors have the same number of levels in both settings.

- The number of response variables: P = 2,4,8, where the number of active variables (under the alternative hypotesis) is 2 when P=2, is 2 when P=4 and is 4 when P=8.

- The dependence and independence among response variables. In case of independence, the variance/covariance matrix is the identity matrix $\boldsymbol{I}_P$ where $\sigma_{rc} = 0$, $\forall r, c = 1, \ldots, P$, $r \neq c$. In case of dependence, the variance/covariance matrix is $\boldsymbol{\Sigma}_P$ where $\sigma_{rc} = 0.5$, $\forall r, c = 1, \ldots, P$, $r \neq c$.

- The number of replicates: $n = 3, 5$

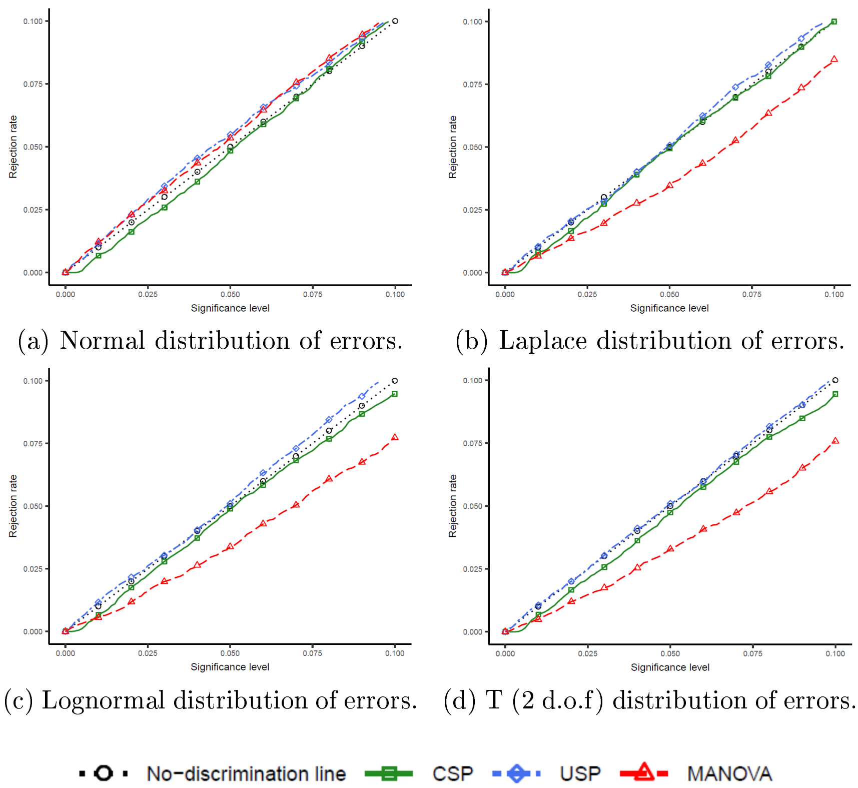
It is well known that the number of replicates affects positively the power of the tests as it increases. Studying the performance in case of low number of replicates reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. The SP tests (CSP, USP) combined with the three combining functions (Fisher, Liptak and Tippet) of NPC methods will be investigated in the 24 settings defined as combination of the varying parameters, and will be compared with the MANOVA test along the four distribution functions (normal, Laplace, lognormal and student's t). Furthermore, some simulations with 100 and 50 response variables (50 and 25 active variables respectively) are run with covariance = 0.5, number of levels = 2 and number of replicates = 5, to investigate the behavior of NPC applied to Synchronized Permutation tests with an high number of response variables.

All simulations are performed in R (version 3.4.0; R Development Core Team (2017)). The number of simulations is $n_{sim} = 10000$, and the number of permutations for CSP and USP is $n_{perm} = 2000$.

6.5.1 Results

Graphical representation of results allows the evaluation of performances of the tests and their comparison in a clear and effective way. In this section main results of the simulation study are presented. The graphs have been obtained plotting the rejection rate of the test (y axis) versus the significance level (x axis). The objective is to compare the performance of the NPC combining functions applied to the permutation tests in the range of interest of significance level $0 \leq \alpha \leq 0.1$. Because of the simmetry of the simulation design, we have same results for factor A and factor B.

In Figure 6.1 it is clear that in case of non normal distribution of errors, the MANOVA test does not respect the α level under the null hypothesis unlike the NPC tests. The curves can be thought as the cumulative distribution functions (y axis) of the related p-values (x axis). There is a discrepancy between the MANOVA curve and the line of no-discrimination when $H_0^{(A)}$ is true, while in general NPC tests' curves are very close to the hypothetical



(a) Normal distribution of errors. (b) Laplace distribution of errors.

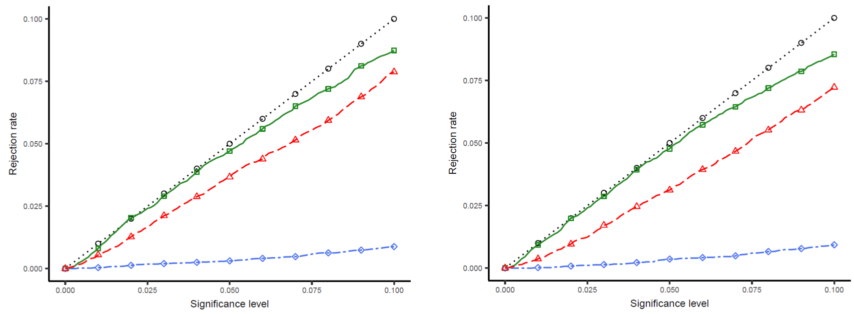
(c) Lognormal distribution of errors. (d) T (2 d.o.f) distribution of errors.

Figure 6.1: Behaviour of tests under null hypothesis. Rejection rate at different values of significance level α (x axis). NPC function = Fisher; factor A; number of responses = 4; number of levels = 2; covariance = 0; number of replicates = 5.

continuous uniform distribution with every combining function for the main factor (factor A).

The respect of the α level under the null hypothesis in the analysis of interaction effect is more challenging for all the considered tests. Under non normality MANOVA shows a clear departure from no-discrimination line. The multivariate combination of USP test reveals to be unreliable as well. The most performing test for the interaction effect in the model assumed is the Liptak combination of CSP test (Figure 6.2).

In Section 6.2 the issue of the minimum achievable significance level related to the cardinality of the univariate permutation tests is shown both for CSP and USP. The curves of NPC of Synchronized Permutations reflect the same issue with an initial plateau with rejection rate = 0. A clear example is given by the case of CSP and a data set with $(I, J) = (2, 2)$ levels for the



(a) Lognormal distribution of errors. (b) T (2 d.o.f) distribution of errors.

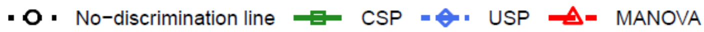
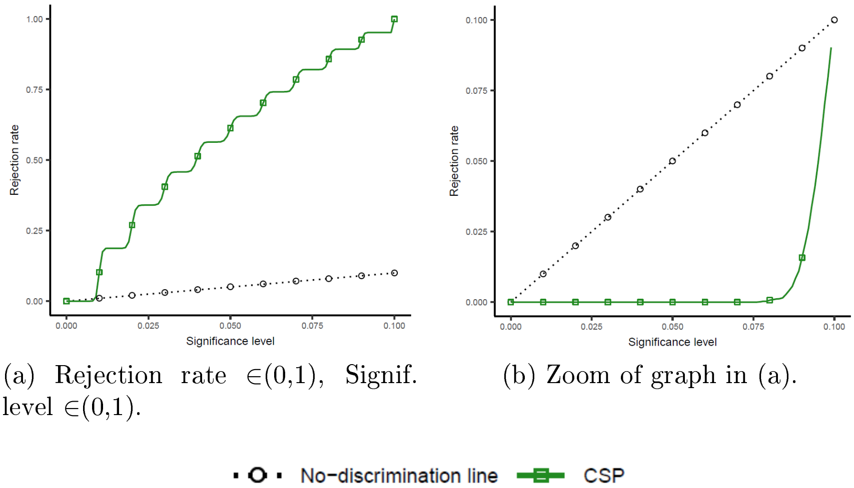


Figure 6.2: Behaviour of tests under null hypothesis. Rejection rate at different values of significance level α (x axis). NPC function = Liptak; interaction AB; number of responses = 4; number of levels = 2; covariance = 0; number of replicates = 5.

factors A and B respectively and $n = 3$ replicates. The cardinality of the univariate test is $C_{csp} = \binom{2n}{n} = \binom{6}{3} = 20$ and the univariate minimum achievable significance error is $\alpha_{min} = 2 \times (C_{csp})^{-1} = 0.1$. The set of exact p-values that can result under such conditions is a set of 10 values at step of 0.1: $S_{p-values} = \{0.1, \ldots, 1\}$ for each univariate test. The application of a combining function gives a set of 10 values for the statistic T'' used to compute the p-value of the second-order test $\hat{\lambda}''_{\phi}$. The cumulative distribution function of the p-values is shown in Figure 6.3 (a) where we can recognize 10 steps. Recall that we are using a CMC procedure with a number of permutations large enough so that the length of the steps is quite regular thanks to the fact that the likelihood of the values of the statistic T'' is the same. A zoom in Figure 6.3 (b) shows clearly the initial plateau with rejection rate = 0. The NPC tests are affected by the minimum achievable significance level of the permutation test they are applied to. All the graphs in this section show a small initial plateau. This fact reveals to be a shortcoming only in the case of CSP and 3 replicates considering the usual significance levels $\alpha = 0.01$ and $\alpha = 0.05$.

The capability of NPC applied to SP tests to detect the effect of the main factor under H_1 in the assumed model is in general good. Simulation results show that NPC applied to USP and CSP gives high values of power (rejection rate) both with independent and dependent response variables, and both with low number and high number of response variables compared



(a) Rejection rate ∈(0,1), Signif. level ∈(0,1).

(b) Zoom of graph in (a).

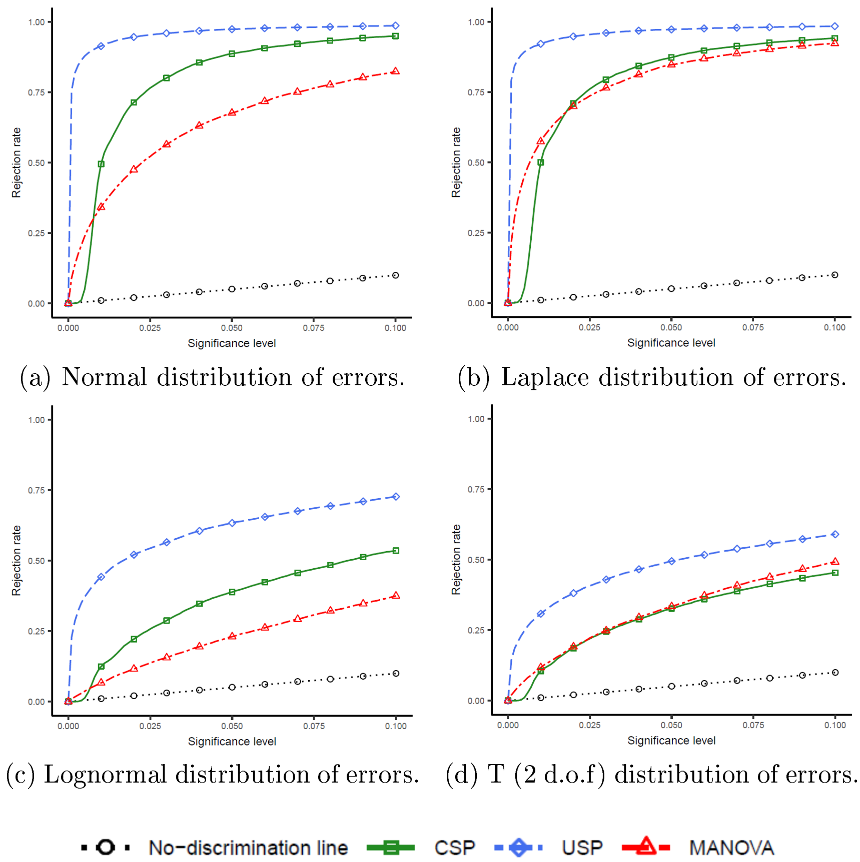
Figure 6.3: Effect of cardinality of Synchronized Permutations on minimum significance level of NPC. Rejection rate at different values of significance level α (x axis). NPC function = Fisher; factor A; number of responses = 4; distribution of errors = T (2 d.o.f) ; number of levels = 2; covariance = 0.5; number of replicates = 3.

to MANOVA (Figures 6.4, 6.5). In general NPC applied to USP performs better with all distribution of errors, while NPC of CSP and MANOVA have in some cases similar performances with Laplace and student's t distribution of errors. NPC tests show a better performance compared to MANOVA even in case of normal distribution of errors with the power that can be even more than double at $\alpha = 0.05$ significance level (Figure 6.4).

In the interaction analysis, the observed rejection rate under H_1 is lower than in the main factor analysis (Figure 6.6). This result is consistent with the model used to generate data set for simulation study where interaction is given by the product of the effects of the factors, and with the values of the factor level effects. In general, the NPC of CSP performs better than MANOVA and USP, both with independent and dependent response variables, and both with low number and high number of response variables. The NPC of USP shows the lower power. In some cases with two levels of factor MANOVA and NPC applied to CSP have similar rejection rate, with MANOVA performing better with Laplace distribution of errors.

The increase of number of levels of factors A and B from $(I, J) = (2, 2)$ to $(I, J) = (3, 3)$ has a positive effect on the rejection rate under H_1 (Figure 6.7) even if the maximum $\delta = 1$ between factors is constant.

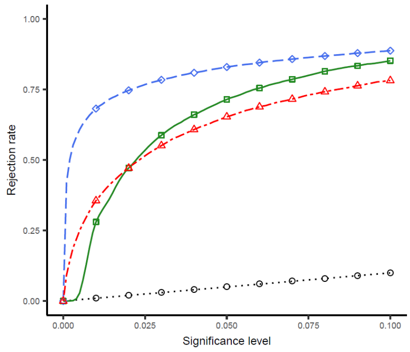
The three combining functions Fisher, Liptak and Tippet show in general



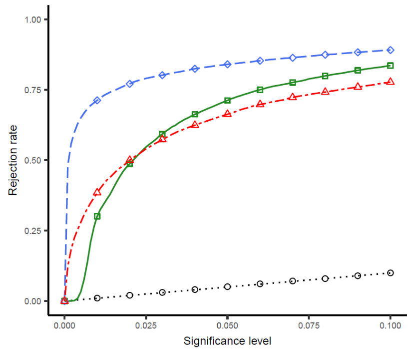
(a) Normal distribution of errors. (b) Laplace distribution of errors.

(c) Lognormal distribution of errors. (d) T (2 d.o.f) distribution of errors.

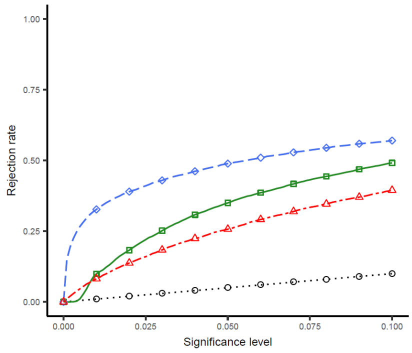
Figure 6.4: Comparison of Non-Parametric methods and MANOVA. Rejection rate at different values of significance level α (x axis). NPC function = Fisher; factor A; number of responses = 8; number of levels = 3; covariance = 0; number of replicates = 5.



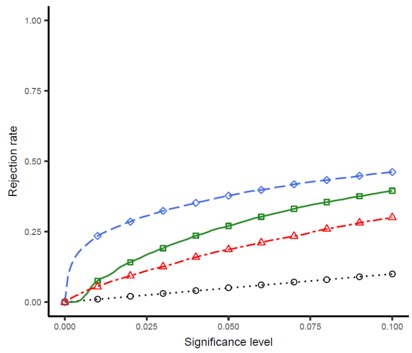
(a) Normal distribution of errors.



(b) Laplace distribution of errors.



(c) Lognormal distribution of errors.



(d) T (2 d.o.f) distribution of errors.

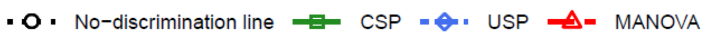
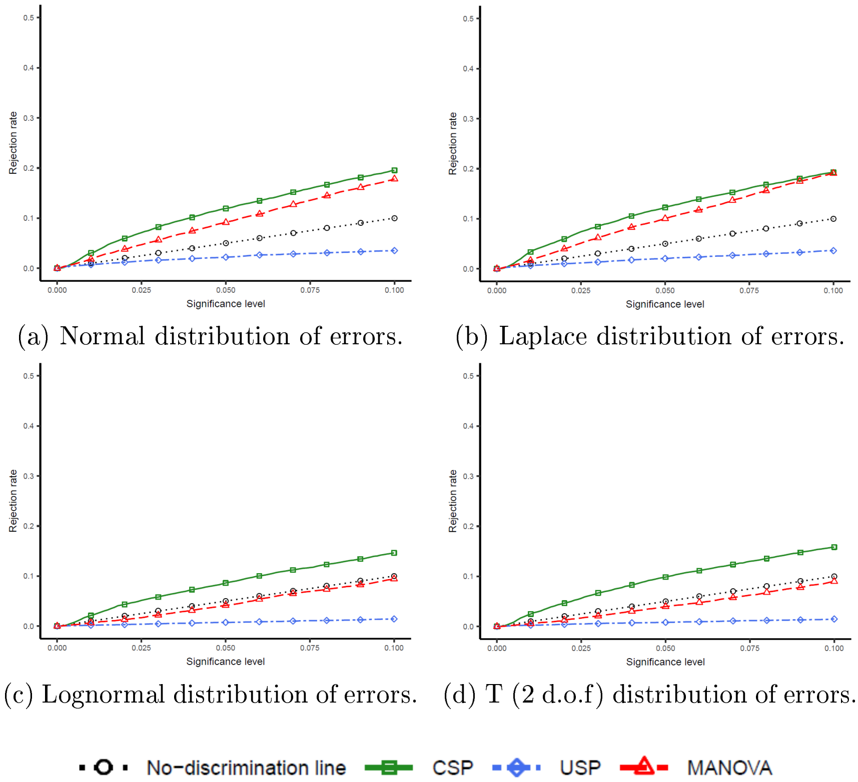


Figure 6.5: Comparison of Non-Parametric methods and MANOVA. Rejection rate at different values of significance level α (x axis). NPC function = Fisher; factor A; number of responses = 2; number of levels = 3; covariance = 0.5; number of replicates = 5.



(a) Normal distribution of errors. (b) Laplace distribution of errors.

(c) Lognormal distribution of errors. (d) T (2 d.o.f) distribution of errors.

Figure 6.6: Comparison of Non-Parametric methods and MANOVA. Rejection rate at different values of significance level α (x axis). NPC function = Liptak; interaction AB; number of responses = 4; number of levels = 3; covariance = 0.5; number of replicates = 5.

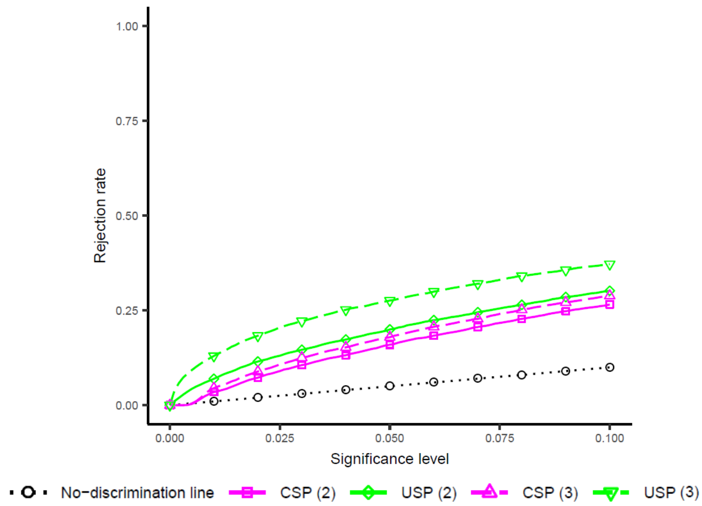
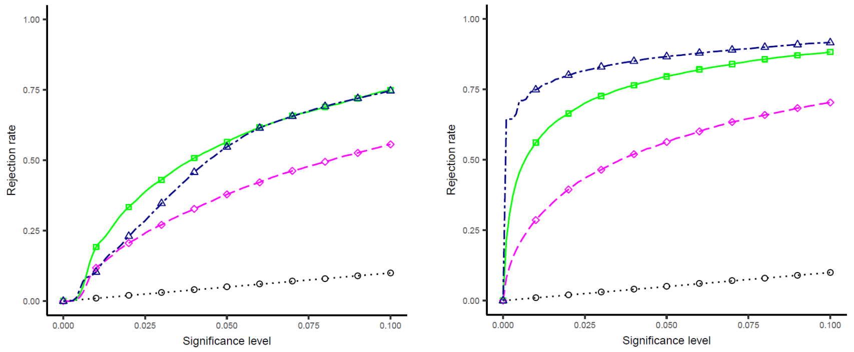


Figure 6.7: Performance at different number of levels, 2 and 3. Rejection rate at different values of significance level α (x axis). NPC function = Fisher; factor A; distribution of errors = T (2 d.o.f); number of responses = 4; covariance = 0.5; number of replicates = 5.

good performances. Nevertheless, Fisher and Tippet functions give higher rejection rate compared to Liptak, with Tippet function's curve showing small steps (Figure 6.8).

An increase in rejection rate can be observed when the number of response variables increases with fixed number of observed units, meaning an higher power of the test in detecting a factor effect under H_1 (Figure 6.9, 6.10). This phenomenon is known as *finite sample consistency* and refers to a peculiar property of multivariate combination-based inferences: the power of NPC tests for any added variable monotonically increases if the variable makes larger noncentrality parameter of the underlying population distribution [62, 66]. The positive effect on the power of the test that can be obtained adding response variables can be strategically exploited considering that in many real problems it could be easier to collect more information on a single experimental unit than adding a new unit to the experimental design [16]. The effect of the increase of response variables while keeping constant the number of observed units couldn't be investigated for MANOVA test because of the problem of the loss of degrees of freedom that does not allow to apply MANOVA test when the number of response variables is larger than the sample size.



(a) Constrained Synchronized Permutation.

(b) Unconstrained Synchronized Permutation.

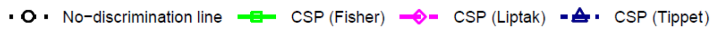
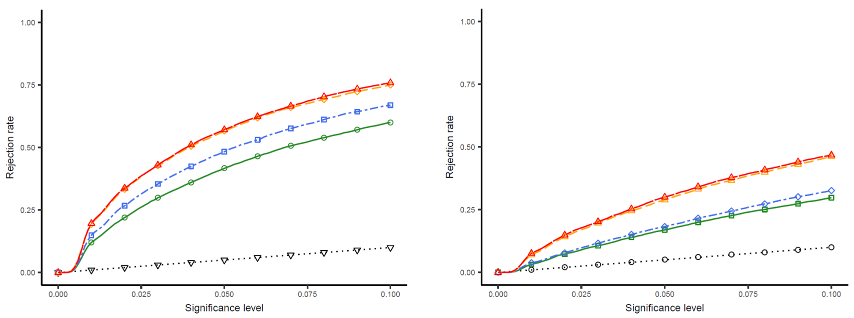


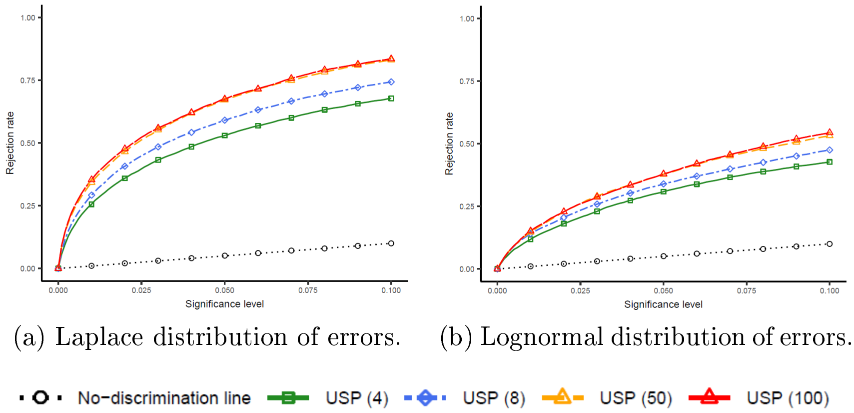
Figure 6.8: Comparison of NPC functions. Rejection rate at different values of significance level α (x axis). Factor A; distribution of errors = Laplace; number of responses = 8; number of levels = 3; covariance = 0.5; number of replicates = 5.



(a) Laplace distribution of errors.

(b) Lognormal distribution of errors.

Figure 6.9: Effect of increasing number of responses with CSP. Rejection rate at different values of significance level α (x axis). NPC function = Fisher, factor A; number of levels = 2; covariance = 0.5; number of replicates = 5.



(a) Laplace distribution of errors. (b) Lognormal distribution of errors.

Figure 6.10: Effect of increasing number of responses with USP. Rejection rate at different values of significance level α (x axis). NPC function = Fisher, factor A; number of levels = 2; covariance = 0.5; number of replicates = 5.

6.6 Application to Innovation of a Production Line

A real case study is useful to highlight the benefits of the adoption of the herein presented nonparametric approach in industrial experiments with a small sample size and non-normal data distribution. An industrial experiment according to a two-way two-levels design in the engineering field provides an example of the analysis performed using NPC combined with Synchronized Permutations on a dataset with two responses.

The production system of plastic thermoformed packaging is complex, and it is controlled by several factors [68]. In order to innovate the system, the impact of two factors and of their interaction has to be assessed for values of levels outside their usual range. One factor is the temperature of the process (factor A) and the other factor is the pace of production line (factor B). The packaging is composed by two separate and different-in-shape chambers. The evaluation of the strength of the packaging is done observing the pressure needed to break the packaging by a burst test [bar]. Each chamber is tested separately, so there are two response variables. Five replicates for each factor level combination and for each response variable have been tested. Data collected violate assumption of normality (Table 6.1), and they reveal heteroscedasticity based on the different values of control factors (Table 6.2).

MANOVA test is not reliable for the analysis of such data set. The NPC combined with CSP and USP overcome the violation of Manova assump-

Table 6.1: Multivariate normality tests

	Shapiro-Wilk	Henze-Zirkler	Royston
Test statistic	0.908	1.374	11.626
p-value	0.058	0.001	0.003

Table 6.2: Box's M-test for Homogeneity of Covariance Matrices

	Chi-Sq (approx.)
Test statistic	24.576
DF	9
p-value	0.003

tions. The results of the test are in Table 6.3 and 6.4. Note that interaction effect has been analized only with NPC applied to CSP because USP doesn't respect the α level under the null hypothesis. According to simulation study results, main factor effect should preferably be assessed using NPC applied to USP. The null hypothesis is rejected at a significance level $\alpha = 0.05$. Both the factors and their interaction have a significant impact on the final product. A further investigation allowed to find the setting for the optimal strength of the packaging.

6.7 Conclusions

The novel nonparametric approach in which I suggest the application of NonParametric Combination to Synchronized Permutation to analyze a multivariate two-way factorial design reveals to be a good instrument for inferential statistics when assumptions of MANOVA are violated. Simulation

Table 6.3: Constrained Synchronized Permutation (CSP): p-values of the NPC tests

	Fisher	Liptak	Tippet
Temperature	2.60e-02	3.50e-02	2.70e-02
Cycles per minute	1.50e-02	3.30e-02	7.00e-03
Interaction	7.00e-03	7.00e-03	7.00e-03

Table 6.4: Unconstrained Synchronized Permutation (USP): p-values of the NPC tests

	Fisher	Liptak	Tippet
Temperature	6.00e-03	1.65e-02	4.00e-03
Cycles per minute	1.50e-03	1.50e-03	1.50e-03

results show that NPC applied to USP and CSP gives high values of power (rejection rate) under alternative hypothesis H_1 both with independent and dependent response variables, and both with low number and high number of response variables compared to MANOVA. In general NPC applied to USP performs better for the main factor analysis with all distribution of errors compared to NPC of CSP and MANOVA. Its power varies under the conditions it has been tested in the simulation study, and it has been observed to be higher than 75% at a significance level $\alpha = 0.05$ in many cases. For the interaction analysis we recommend the adoption of NPC of CSP with the Liptak combining function because of the higher adherence of the test to the α nominal level. The Fisher combining function, also referred to as *omnibus*, is in general preferable to the Tippet and the Liptak ones in the main factor analysis.

A great advantage given by the adoption of these tests is that they well perform with small sample size. This reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. In case of $n = 3$ replicates we recommend the use of NPC of USP for main factor analysis because of the shortcomings of the minimum achievable significance level related to the cardinality of the univariate CSP test. The increase of sample size has in general an evident positive effect on the power of NPC of SP tests. Futhermore the power of the test is improved by the increase of number of factor levels with the the factor effect fixed.

Last, there is an important property of NPC of SP tests that can be exploited to increase their power: the finite sample consistency. Indeed, an increase in rejection rate can be observed under alternative hypothesis H_1 when the number of response variables increases with fixed number of observed units. This could lead to a strategical benefit considering that in many real problems it could be easier to collect more information on a single experimental unit than adding a new unit to the experimental design.

Chapter 7

Discussion and Conclusions

The research conducted as PhD student and results presented in previous chapters of this dissertation let me be on the same side of those scholars that in recent scientific literature stand in general in favour of the adoption of Design of Experiment for innovation, and in particular in favour of the adoption of DOE for production process innovation. Each step of the research was conducted with the aim of answering to the research questions that emerged from the systematic literature review. Nevertheless, a fil rouge ties together the case study by means of which I answered to the first and second research questions and the simulations studies by means of which I faced the third research question. The fil rouge consists in the challenges that I had to face in the analysis of the data set of the case study. This is typical when a research is conducted. You move on adapting step by step to the results and discoveries that you make, and new scenarios and opportunities for the research show up.

Main results and conclusions of the research are herein summarized following the same structure of the dissertation.

An experimental strategy on innovation of thermoforming production process has been developed. Design of Experiment (DOE) techniques were used in designing and analyzing all the phases of the strategy. DOE enhanced innovation capability allowing reduction of systematic errors and distortions, full exploration of factorial space, and reduction of number of tests. The experimental strategy allows selection of material and correlation of control factors levels to packaging performance for each tested material. Traditional approach to production control in thermoforming process was challenged. DoE allowed to identify and overcome the mismatch between control factors in laboratory and in production line.

The qualitative research performed during the adoption of DOE for the development of a new packaging production process allowed to highlight the impact of DOE on innovation process management in the specific case analyzed. The introduction of DOE brought a new perspective on the innovation process. The new perspective was achieved even by asking fundamental questions and challenging basic assumptions. But, the most important thing was that it was a common perspective among the members of the team. The vision of the process became broad and there were a clear interpretation of correlation between test results and control factors. The latter was not possible before the adoption of DOE. The common goal and the common understanding of the objective of the experimentation lead people to work as one team. The barriers between the teams of experimenters were removed thanks to the end to end involvement. DOE became a common language. It facilitated decision making by the team avoiding conflicts due to subjective opinions. The way results of analysis are shown enhances communication. The quality and effectiveness of information were increased. This creates a fertile field to knowledge building, by means even of the improved system of coding experiments and recording results. The learnings could be in some way generalized beyond the specific case along five dimensions that have a more general value in the managerial field. Namely: decision making, integration, communication, time and cost, and knowledge.

There is anyway one risk that I perceived. The risk is that someone performs unnecessary tests only to seek for the "DOE blessing" in order to not expose himself to criticism. This behavior is contrary to the reason why DOE should be adopted and the benefits it can give. The experimental design should always move first steps from the definition of the problem and not jumping to the experimental phase just to show some data analysis.

A simulation study allowed to assess the power of some selected nonparametric methods by analyzing the same data set. Data have been generated using a linear additive model for a two-way two-levels design with interaction given by the product of factor level effects. Such model is common for practitioners in industrial experimentation. The study revealed that certain methods of analysis perform better than others depending on the dataset and on the objective of the analysis. As a consequence, there does not emerge a unique approach in the design phase of the experiment, but various aspects have to be taken into account simultaneously. The three dimensions (factor effect, standard deviation, number of replicates) along which the investigation has been conducted have impacts on the power of the tests. Furthermore, the study allowed to bring out some interesting information.

Concerning the balanced case, Aligned Rank Transform (ART) test is an overall well performing test both for the main factor and the interaction analysis. It has superior power in almost all the settings, and it maintains the α level well. Wald-Type Prmutation (WTP) test is a good test as well both for main factor and interaction analysis. These findings corroborate with previous studies ([25, 61]). ANOVA-Type Statistic (ATS) test's performance is at an average level for main factor analysis, showing slightly lower power in the heteroscedastic case and resulting conservative in some cases. In the interaction analysis ATS is performing worse relative to the other tests. In the main factor analysis, Unconstrained Synchronized Permutation (USP) is performing well in most of the situations and its power is similar to that of WTP. In the interaction analysis, USP is performing best compared to the other tests in all the cases, but it doesnot maintain the nominal α level. The liberal behavior is a limitation of this test because of the tradeoff between power and type I error. Constrained Synchronized Permutation (CSP) does not reveal any liberal behavior in the interaction analysis and its power is similar to that of WTP. In agreement to results of Hahn et al. [25], in the main factor analysis CSP has, in general, slightly lower power for Laplace and lognormal distribution relative to the other tests. In the main factor analysis, there is no influence on the power of the tests due to the other factor's level. The results for CSP and USP are also in agreement with results of a previous study ([3]).

In the unbalanced case, WTP and ATS tests are the only ones that appear to be reliable tests in all the scenarios considered. Indeed, they control the α level and maintain the power level when the number of replicates is switched between the factor level combinations. CSP and USP tests can be used due to the *fixed weight* approach only when $n_{11} = n_{12}$ and $n_{21} = n_{22}$, but they do not control the α level, resulting in conservative decisions in the so-called *positive pairing* heteroscedastic setting and liberal decision in the *negative pairing* for the main factor analysis as well as the interaction analysis, unlike the ART test. In the homoscedastic case, ART test performs better than WTP and ATS, and its power varies when the number of replicates is switched between the factor level combination. In the main factor analysis, ART test reveals an influence on its power due to the level of the other factor in some configurations.

The application of NonParametric Combination (NPC) to Synchronized Permutation (SP) to analyze a multivariate two-way factorial design revealed to be a good instrument for inferential statistics when assumptions of MANOVA are violated. Simulation results show that NPC applied to USP and CSP gives high values of power (rejection rate) under alternative hypothesis H_1 both with independent and dependent response variables, and both with low number and high number of response variables compared to MANOVA. In general NPC applied to USP performs better for the main factor analysis with all distribution of errors compared to NPC of CSP and MANOVA. Its power varies under the conditions it has been tested in the simulation study, and it has been observed to be higher than 75% at a significance level $\alpha = 0.05$ in many cases. For the interaction analysis the adoption of NPC of CSP with the Liptak combining function has to be recommended because of the higher adherence of the test to the α nominal level. The Fisher combining function, also referred to as *omnibus*, is in general preferable to the Tippet and the Liptak ones in the main factor analysis. A great advantage given by the adoption of these tests is that they well perform with small sample size. This reflects the frequent needs of practitioners in the industrial environment where there are constraints or limited resources for the experimental design. In case of $n = 3$ replicates the use of NPC of USP for main factor analysis has to be recommended because of the shortcomings of the minimum achievable significance level related to the cardinality of the univariate CSP test. The increase of sample size has in general an evident positive effect on the power of NPC of SP tests. Futhermore the power of the test is improved by the increase of number of factor levels with the factor effect fixed. At last, there is an important property of NPC of SP tests that can be exploited to increase their power: the finite sample consistency. Indeed, an increase in rejection rate can be observed under alternative hypothesis H_1 when the number of response variables increases with fixed number of observed units. This could lead to a strategical benefit considering that in many real problems it could be easier to collect more information on a single experimental unit than adding a new unit to the experimental design. Properties of this multivariate test make of it a useful instrument when using DOE to innovate a production process and some specific conditions are verified.

Bibliography

[1] J.M. Anderson and P.J. Whitcomb. "Practical aspects for designing statistically optimal experiments". In: *Journal of Statistical Science and Application* 2.3 (2014), pp. 85–92.

[2] Rosa Arboretti et al. "Multivariate small sample tests for two-way designs with applications to industrial statistics". In: *Statistical Papers* (2018), pp. 1–21.

[3] D. Basso et al. *Permutation Tests for Stochastic Ordering and ANOVA: theory and applications with R*. New York, NY: Springer, 2009.

[4] Dario Basso, Marco Chiarandini, and Luigi Salmaso. "Synchronized permutation tests in replicated I× J designs". In: *Journal of Statistical Planning and Inference* 137.8 (2007), pp. 2564–2578.

[5] Arne C Bathke and Solomon W Harrar. "Nonparametric methods in multivariate factorial designs for large number of factor levels". In: *Journal of Statistical planning and Inference* 138.3 (2008), pp. 588–610.

[6] Arne C Bathke, Solomon W Harrar, and Laurence V Madden. "How to compare small multivariate samples using nonparametric tests". In: *Computational Statistics & Data Analysis* 52.11 (2008), pp. 4951–4965.

[7] G.A. Berti et al. "Response surface modelling of thermo-mechanical fatigue in hot forging". In: *Statistica* 66.2 (2006), pp. 171–182.

[8] Søren Bisgaard. "The future of quality technology: From a manufacturing to a knowledge economy & from defects to innovations". In: *Quality Engineering* 24.1 (2012), pp. 30–36.

[9] GEORGE EP BOX. "George's column". In: *Quality Engineering* 2.4 (1990), pp. 365–69.

[10] George EP Box and William H Woodall. "Innovation, quality engineering, and statistics". In: *Quality Engineering* 24.1 (2012), pp. 20–29.

[11] Edgar Brunner, Holger Dette, and Axel Munk. "Box-Type Approximations in Nonparametric Factorial Designs". In: *Journal of the American Statistical Association* 92.440 (1997), pp. 1494–1502.

[12] Edgar Brunner, Ullrich Munzel, and Madan L Puri. "The multivariate nonparametric Behrens–Fisher problem". In: *Journal of Statistical Planning and Inference* 108.1-2 (2002), pp. 37–53.

[13] T.I. Butler and B.A. Morris. "PE-Based Multilayer Film Structures". In: *Plastic Films in Food Packaging: Materials, Technology and Applications*. Ed. by Sina Ebnesajjad. New York, NY: Elsevier Inc., 2012, pp. 21–52.

[14] William J Conover and Ronald L Iman. "Rank transformations as a bridge between parametric and nonparametric statistics". In: *The American Statistician* 35.3 (1981), pp. 124–129.

[15] L. Corain and L. Salmaso. "Nonparametric permutation and combination-based multivariate control charts with applications in microelectronics". In: *Applied Stochastic Models in Business and Industry* 29.4 (2013), pp. 334–349.

[16] Livio Corain and Luigi Salmaso. "Improving power of multivariate combination-based permutation tests". In: *Statistics and Computing* 25.2 (2015), pp. 203–214.

[17] W Edwards Deming. "The new economics for industry, government, education. Cambridge, Massachusetts: Massachusetts Institute of Technology". In: *Center for Advanced Engineering Study* (1993).

[18] N.R. Draper and F. Pukelsheim. "An overview of design of experiments". In: *Statistical Papers* 37.1 (1996), pp. 1–32. DOI: `10.1007/BF02926157`.

[19] E. Edgington and P. Onghena. *Randomization Tests, Fourth Edition*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 2007. ISBN: 9781584885894.

[20] Sarah Friedrich, Frank Konietschke, and Markus Pauly. "GFD: An R Package for the Analysis of General Factorial Designs". In: *Journal of Statistical Software, Code Snippets* 79.1 (2017), pp. 1–18.

[21] Sarah Friedrich, Frank Konietschke, and Markus Pauly. *GFD: Tests for General Factorial Designs*. R package version 0.2.4. 2017. URL: `https://cran.r-project.org/web/packages/GFD/index.html`.

[22] M. Gloet and D. Samson. “Capturing value through knowledge and innovation management: Comparisons across the manufacturing and services sectors”. In: vol. 2015-March. 2015, pp. 3730–3739.

[23] Phillip I. Good. *Permutation, Parametric, and Bootstrap Tests of Hypotheses.* Springer Series in Statistics. Springer, 2005.

[24] Constantinos Goutis, George Casella, and Martin T Wells. “Assessing evidence in multiple hypotheses”. In: *Journal of the American Statistical Association* 91.435 (1996), pp. 1268–1277.

[25] S. Hahn, F. Konietschke, and L. Salmaso. “A Comparison of Efficient Permutation Tests for Unbalanced Anova in Two by Two Designs and Their Behavior Under Heteroscedasticity”. In: *Springer Proceedings in Mathematics and Statistics.* Vol. 114. 2014, pp. 257–269.

[26] S. Hahn and L. Salmaso. “A comparison of different synchronized permutation approaches to testing effects in two-level two-factor unbalanced ANOVA designs”. In: *Statistical Papers* 58.1 (2017), pp. 123–146.

[27] A Hargadon. “How Breakthrough Happen: the Surprising Truth about How Companies Innovate”. In: *Boston: MA* (2003).

[28] Solomon W Harrar and Arne C Bathke. “A modified two-factor multivariate analysis of variance: asymptotics and small sample approximations”. In: *Annals of the Institute of Statistical Mathematics* 64.1 (2012), pp. 135–165.

[29] Solomon W Harrar and Arne C Bathke. “A nonparametric version of the Bartlett-Nanda-Pillai multivariate test. Asymptotics, approximations, and applications”. In: *American Journal of Mathematical and Management Sciences* 28.3-4 (2008), pp. 309–335.

[30] Solomon W Harrar and Arne C Bathke. “Nonparametric methods for unbalanced multivariate data and many factor levels”. In: *Journal of Multivariate Analysis* 99.8 (2008), pp. 1635–1664.

[31] James J Higgins, R Clifford Blair, and Suleiman Tashtoush. “The aligned rank transform procedure”. In: (1990).

[32] James J Higgins and S Tashtoush. “An aligned rank transform test for interaction”. In: *Nonlinear World* 1.2 (1994), pp. 201–211.

[33] Brian Hindo. “3M’s innovation crisis: How Six Sigma almost smothered its idea culture”. In: *Business Week* (2007), pp. 8–14.

[34] Wassily Hoeffding. "The large-sample power of tests based on permutations of observations". In: *The Annals of Mathematical Statistics* (1952), pp. 169–192.

[35] Roger W Hoerl and Ron Snee. "Statistical thinking and methods in quality improvement: a look to the future". In: *Quality engineering* 22.3 (2010), pp. 119–129.

[36] Myles Hollander and Douglas Wolfe. *Nonparametric Statistical Methods.* New York: Wiley, 2013.

[37] T. Iwasaki, W. Takarada, and T. Kikutani. "Influence of processing conditions on heat sealing behavior and resultant heat seal strength for peelable heat sealing of multilayered polyethylene films". In: *Journal of Polymer Engineering* 36.9 (2016), pp. 909–916. DOI: `10.1515/polyeng-2015-0383`.

[38] R. Jarrett. "Does theory work in practice? Two case studies". In: *Quality Engineering* 29.1 (2017), pp. 141–159.

[39] Willis Jensen et al. "Statistics to Facilitate Innovation*: A Panel Discussion". In: *Quality Engineering* 24.1 (2012), pp. 2–19.

[40] R.T. Johnson, D.C. Montgomery, and B.A. Jones. "An expository paper on optimal design". In: *Quality Engineering* 23.3 (2011), pp. 287–301. DOI: `10.1080/08982112.2011.576203`.

[41] S Johnson. *Where Good Ideas Come from: The Natural History of Innovation. 2010 New York.*

[42] Matthew Kay and Jacob O. Wobbrock. *ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs.* R package version 0.10.4. 2016. URL: `https://github.com/mjskay/ARTool`.

[43] K. Khaleel et al. "Using design of experiment to predict concert compressive strength using fly ash and GGBFS as cement alternative". In: *Proceedings of the International Conference on Industrial Engineering and Operations Management.* 2017, pp. 741–749.

[44] Hussein Khreis et al. "Sensitivity analysis for induction machine manufacturing tolerances: Modeling of electrical parameters deviation". In: *Ecological Vehicles and Renewable Energies (EVER), 2017 Twelfth International Conference on.* IEEE. 2017, pp. 1–9.

[45] William H Kruskal. "A nonparametric test for the several sample problem". In: *The Annals of Mathematical Statistics* (1952), pp. 525–540.

[46] William H Kruskal and W Allen Wallis. "Use of ranks in one-criterion variance analysis". In: *Journal of the American statistical Association* 47.260 (1952), pp. 583–621.

[47] M. Labonte and C. Dubois. "Optimization of molding conditions of a plug-assisted thermoformed thin containers in a high speed and volume production context". In: *Annual Technical Conference - ANTEC, Conference Proceedings*. Vol. 3. 2011, pp. 2515–2519.

[48] Daniël Lakens. "Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs". In: *Frontiers in psychology* 4 (2013), p. 863.

[49] Viliam Lendel, Štefan Hittmár, and Eva Siantová. "Identification of the Main Problems in the Management of Innovation Processes and the Draft of Appropriate Recommendations". In: *International Conference on Knowledge Management in Organizations*. Springer. 2015, pp. 221–232.

[50] T Liptak. "On the combination of independent tests". In: *Magyar Tud Akad Mat Kutato Int Kozl* 3 (1958), pp. 171–197.

[51] Q. Liu, X. Chen, and N. Gindy. "Evaluation of superalloy heavy-duty grinding based on multivariate tests". In: *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* 221.9 (2007), pp. 1421–1430.

[52] D Macri and Maria Rita Tagliaventi. *La ricerca qualitativa nelle organizzazioni. Teorie, tecniche, casi.* Carocci, 2000.

[53] V.D. Milovidov. "Management of innovations: How to effectively use the information". In: *Neftyanoe khozyaystvo - Oil Industry* 6 (2015), pp. 10–16.

[54] D.C. Montgomery. *Design and Analysis of Experiments*. Danvers, MA: John Wiley & Sons,Inc., 2000.

[55] Douglas C Montgomery. "Applications of design of experiments in engineering". In: *Quality and Reliability Engineering International* 24.5 (2008), pp. 501–502.

[56] Douglas C Montgomery and Rachel T Silvestrini. "Design of experiments: A key to successful innovation". In: *12th International Workshop on Intelligent Statistical Quality Control, IWISQC 2016*. Helmut Schmidt University. 2016.

[57] R. Morales and M.V. Candal. "Thermoforming process optimization by using the experiment design technique". In: *Annual Technical Conference - ANTEC, Conference Proceedings*. Vol. 5. 2006, pp. 2641–2645.

[58] Ullrich Munzel and Edgar Brunner. "Nonparametric methods in multivariate factorial designs". In: *Journal of Statistical Planning and Inference* 88.1 (2000), pp. 117–132.

[59] I.B. Onukogu and M.P. Iwundu. "A combinatorial procedure for constructing D-optimal exact designs". In: *Statistica* 67.4 (2007), pp. 415–423.

[60] M. Ota, Y. Hazama, and D. Samson. "Japanese innovation processes". In: *International Journal of Operations and Production Management* 33.3 (2013), pp. 275–295.

[61] M. Pauly, E. Brunner, and F. Konietschke. "Asymptotic permutation tests in general factorial designs". In: *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 77.2 (2015), pp. 461–473.

[62] F. Pesarin and L. Salmaso. *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley Series in Probability and Statistics. Wiley, 2010. ISBN: 9780470689523.

[63] Fortunato Pesarin. "A resampling procedure for nonparametric combination of several dependent tests". In: *Journal of the Italian Statistical Society* 1.1 (1992), pp. 87–101.

[64] Fortunato Pesarin. *Multivariate permutation tests: with applications in biostatistics*. Vol. 240. Wiley Chichester, 2001.

[65] Fortunato Pesarin. *Permutation testing of multidimensional hypotheses by nonparametric combination of dependent tests*. Cleup, 1999.

[66] Fortunato Pesarin and Luigi Salmaso. "Finite-sample consistency of combination-based permutation tests with application to repeated measures designs". In: *Journal of Nonparametric Statistics* 22.5 (2010), pp. 669–684.

[67] Michael E Porter and Scott Stern. "National innovative capacity". In: *The global competitiveness report* 2002 (2001), pp. 102–118.

[68] Fabrizio Ronchi et al. "Optimal Designs to Develop and Support an Experimental Strategy on Innovation of Thermoforming Production Process". In: *Statistica* 77.2 (2017), p. 109.

[69] L. Salmaso. "Synchronized permutation tests in 2k factorial designs". In: *Communications in Statistics - Theory and Methods* 32.7 (2003), pp. 1419–1437.

[70] M. Sameh Ibrahim, M.A.R. Mansour, and M. Abed. “Improving the quality of thermoforming process using a proposed analysis algorithm (a case study)”. In: *Applied Mechanics and Materials* 110-116 (2012), pp. 1466–1476. DOI: 10.4028/www.scientific.net/AMM.110-116.1466.

[71] D. Samson, M. Gloet, and P. Singh. “SYSTEMATIC INNOVATION CAPABILITY: EVIDENCE from CASE STUDIES and A LARGE SURVEY”. In: *International Journal of Innovation Management* 21.7 (2017).

[72] Kusalin Sangnuan and Wimalin S Laosiritaworn. “Determining the optimal parameter of coordinate measuring machine with design of experiment”. In: *MATEC Web of Conferences*. Vol. 68. EDP Sciences. 2016, p. 06009.

[73] J.R. Smith and U.K. Vaidya. “Processing optimization of deformed plain woven thermoplastic composites”. In: *Applied Composite Materials* 20.6 (2013), pp. 1265–1272. DOI: 10.1007/s10443-013-9333-8.

[74] Z. Stacho et al. “The organizational culture as a support of innovation processes’ management: A case study”. In: *International Journal for Quality Research* 10.4 (2016), pp. 769–784.

[75] A. Toillier et al. “Understanding the contribution of research to collective innovation through the exploration of capacity development mechanisms”. In: *Cahiers Agricultures* 27.1 (2018).

[76] David Tranfield, David Denyer, and Palminder Smart. “Towards a methodology for developing evidence-informed management knowledge by means of systematic review”. In: *British journal of management* 14.3 (2003), pp. 207–222.

[77] J. Van Lancker, E. Wauters, and G. Van Huylenbroeck. “Managing innovation in the bioeconomy: An open innovation perspective”. In: *Biomass and Bioenergy* 90 (2016), pp. 60–69.

[78] C. Wang and Q. Hu. “Knowledge sharing in supply chain networks: Effects of collaborative innovation activities and capability on innovation performance”. In: *Technovation* (2017).

[79] Hajo Wiemer et al. “A Holistic and DoE-based Approach to Developing and Putting into Operation Complex Manufacturing Process Chains of Composite Components”. In: *Procedia CIRP* 66 (2017), pp. 147–152.

[80] Rand R Wilcox. *Introduction to robust estimation and hypothesis testing*. Academic press, 2011.

[81] Jacob O Wobbrock et al. "The aligned rank transform for nonparametric factorial analyses using only anova procedures". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2011, pp. 143–146.

[82] K. Yamada et al. "Peelability and morphology of easy-peel films". In: *Annual Technical Conference - ANTEC, Conference Proceedings*. Vol. 2. 2012, pp. 1361–1365.

[83] Q. Zhang et al. "Experiment of integrated fermentation hydrogen production by photosynthetic bacteria cooperating with Enterobacter aerogenes". In: *Nongye Gongcheng Xuebao/Transactions of the Chinese Society of Agricultural Engineering* 33.9 (2017), pp. 243–249.

[84] M. Zhao and L. Lu. "Effect of heat sealing temperature on heat-sealing performance of PET/Al/PE packaging laminated film". In: *Hecheng Shuzhi Ji Suliao/China Synthetic Resin and Plastics* 25.1 (2008), pp. 57–61.

[85] S.N. Zolnikova, L.M. Saparmuradova, and E.G. Kulchikhina. "Management of an enterprise innovative activity". In: *Academy of Strategic Management Journal* 16.Special issue 2 (2017).