UNIVERSITÀ DEGLI STUDI DI GENOVA

PH.D THESIS
XXXIII COURSE

Design and optimization under uncertainty of Energy Systems

Author:
Andrea GIUGNO

Supervisors:
Prof. Alberto TRAVERSO
Prof. Alessandro SORCE

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Università degli studi di Genova
DIME sez. MASET

May 8, 2021

UNIVERSITÀ DEGLI STUDI DI GENOVA

TESI DI DOTTORATO
XXXIII CICLO

Design e Ottimizzazione stocastica di sistemi energetici

Autore:
Andrea GIUGNO

Tutor:
Prof. Alberto TRAVERSO
Prof. Alessandro SORCE

Una tesi presentata in adempimento dei requisiti per il titolo di Dottorato

presso

Università degli studi di Genova
DIME sez. MASET

May 8, 2021

Abstract

In many engineering design and optimisation problems, the presence of uncertainty in data and parameters is a central and critical issue. The analysis and design of advanced complex energy systems is generally performed starting from a single operating condition and assuming a series of design and operating parameters as fixed values. However, many of the variables on which the design is based are subject to uncertainty because they are not determinable with an adequate precision and they can affect both performance and cost. Uncertainties stem naturally from our limitations in measurements, predictions and manufacturing, and we can say that any system used in engineering is subject to some degree of uncertainty. Different fields of engineering use different ways to describe this uncertainty and adopt a variety of techniques to approach the problem. The past decade has seen a significant growth of research and development in uncertainty quantification methods to analyse the propagation of uncertain inputs through the systems. One of the main challenges in this field are identifying sources of uncertainty that potentially affect the outcomes and the efficiency in propagating these uncertainties from the sources to the quantities of interest, especially when there are many sources of uncertainties. Hence, the level of rigor in uncertainty analysis depends on the quality of uncertainty quantification method. The main obstacle of this analysis is often the computational effort, because the representative model is typically highly non-linear and complex. Therefore, it is necessary to have a robust tool that can perform the uncertainty propagation through a non-intrusive approach with as few evaluations as possible.
The primary goal of this work is to show a robust method for uncertainty quantification applied to energy systems. The first step in this direction was made doing a work on the analysis of uncertainties on a recuperator for micro gas turbines, making use of the Monte Carlo and Response Sensitivity Analysis methodologies to perform this study. However, when considering more complex energy systems, one of the main weaknesses of uncertainty quantification methods arises: the extremely high computational effort needed. For this reason, the application of a so-called metamodel was found necessary and useful. This approach was applied to perform a complete analysis under uncertainty of a solid oxide fuel cell hybrid system, starting from the evaluation of the impact of several uncertainties on the system up to a robust design including a multi-objective optimization. The response surfaces have allowed the authors to consider the uncertainties in the system when performing an acceptable number of simulations. These response were then used to perform a Monte Carlo simulation to evaluate the impact of the uncertainties on the monitored outputs, giving an insight on the spread of the resulting probability density functions and so on the outputs which should be considered more carefully during the design phase.
Finally, the analysis of a complex combined cycle with a flue gas condesing heat pump subject to market uncertainties was performed. To consider the uncertainties

in the electrical price, which would impact directly the revenues of the system, a statistical study on the behaviour of such price along the years was performed. From the data obtained it was possible to create a probability density function for each hour of the day which would represent its behaviour, and then those distributions were used to analyze the variability of the system in terms of revenues and emissions.

"As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality."

Albert Einstein

Acknowledgements

Before starting with the actual content of this thesis, I think it is extremely important to express my gratitude to the people who have supported, encouraged and advised me through these years, from the beginning of this journey until the reach of the final goal.
First, I would like to express my deepest gratitude to Prof. Alberto Traverso, for giving me the opportunity to work on this subject and even more for having trusted me and my abilities. I would also like to express deep gratitude to Prof. Alessandro Sorce, who has followed my activities along these years and has given me a great help to overcome the difficulties without saving himself in terms of work and time spent.
Together with them, I also thank Prof. Aristide Fausto Massardo, Prof. Loredana Magistri and Prof. Mario Luigi Ferrari for providing me such a creative and stimulating atmosphere within their research group. They offered me numerous opportunities to grow and challenges with which I could measure myself.
A very big thanks goes to Prof. Shahrokh Shahpar from Rolls Royce for giving me the opportunity to spend one year in his research group and for having taught me so much, also in a field in which I had no expertise. When I look back on that time in Derby, I can see that it was an experience that helped me a lot to grow both personally and professionally.
Thanks also to the staff of Rolls-Royce Fuel Cell Systems and Future Methods for allowing me to work with them.
Thanks to Prof. Contino and Prof. Shahpar for their feedback in reviewing this thesis.
Thanks to all the TPG people for providing such an amazing work environment. I am so happy to have had the opportunity of meeting such people an sharing part of my life with you.
It would be very hard to write the names of all the friends that I would like to thank for being a part of my life, although I have always been fond of the quote "in amicitia vitae" and I think that the friends who have been part of my life know that this part of the acknowledgements goes to them, for the support, the laughs and the time spent together and for "simply" being friends.
Last but not least I want to express my deepest gratitude to my family, without which nothing of this would have been possible, for all the sacrifices they have made for me and for the support they have given which has helped me to achieve my goals, not only professionally speaking.
Finally, thanks to you Anna, for being a very important part of my life, supporting me and sharing your life with me despite the long-distance periods and the difficulties, and having the strength to keep alive this relationship.

Contents

List of Figures

List of Tables

Acronyms

ANOVA Analysis of Variance.

APH Air Pre-Heater.

BLUE Best Linear Unbiased Estimator.

CAPEX CAPital EXpenditure.

CCC Central Composite Circumscribed.

CCF Central Composite Face Centered.

CCGT Combined Cycle Gas-Turbine.

CCI Central Composite Inscribed.

CEPCI Chemical Plant Cost Indexes.

CHEOPE Compact Heat Exchanger Optimisation and Performance Evaluation.

CHP Combined Heat and Power.

CLT Central Limit Theorem.

COP Coefficient Of Performance.

COV Coefficient Of Variance.

DHN District Heating Network.

DoE Design of Experiment.

EOH Effective Operative Hours.

EOL End-Of-Life.

EU European Union.

FAST Fourier Amplitude Sensitivity Test.

FG Flue Gas.

FPH Fuel Pre-Heater.

FPI Fast Probability Integration.

FWH-HT Feed Water Heating High Temperature.

FWH-LT Feed Water Heating Low Temperature.

GHG Green House Gas.

gPC Generalized Polynomial Chaos.

HOB Heat-Only Boiler.

HP Heat Pump.

HRSG Heat Recovery Steam Generator.

HX Heat Exchanger.

IRR Internal Rate of Return.

KPI Key Performance Indicator.

LSM Least Square Method.

MC Monte Carlo.

MGT Micro-Gas Turbine.

MLMC Multi-Level Monte Carlo.

MLMC Sum Squares for Pure Error.

MSPE Mean Square Pure Error.

NIPC Non-Intrusive Polynomial Chaos.

NPV Net Present Value.

O-RSM Optimal Response Surface Methodology.

O& M Operation and Maintenance.

OGB Off Gas Burner.

PBP PayBack Period.

PC Polynomial Chaos.

PDF Probability Density Function.

PI Proportional Integral.

PUN Single National Price.

QMC Quasi-Monte Carlo.

RBDO Reliability-Based Design Optimization.

REC Recuperator.

REF Reformer.

RES Renewable Energy Sources.

RS Response Surface.

RSA Response Sensitivity Analysis.

RSM Response Surface Methodology.

SMR Steam-Methane Reforming.

SOFC Solid Oxide Fuel Cell.

SSE Sum Squares Error.

SSLOF Lack-of-Fit Sum Squares.

SST Sum Squares Total.

STCR Steam-To-Carbon Ratio.

TPG Thermochemical Power Group.

UQ Uncertainty Quantification.

WGS Water-Gas Shift.

WGV Waste-Gate Valve.

Nomenclature

Latin Letters

A	Area $[m^2]$
b	Distance between plates $[m]$
C	Cost [€]
c	Conductivity $[1/\Omega * m]$
c_p	Specific heat capacity $[J/kg * K]$
δ	Fin thickness $[m]$
D_h	Hydraulic diameter $[m]$
f	Fanning friction factor $[-]$
G	Specific flow rate $[kg/m^2s]$
h	convective heat transfer coefficient $[W/m2K]$
H_0	Null hypothesis
H_1	Alternate hypothesis
i	Current $[A]$
k_{ohm}	Ohmic losses coefficient $[-]$
l	Length $[m]$
LHV	Lower heating value $[kJ/mol]$
$\dot{m}$	Mass flow $[kg/s]$
M_x	Main interaction of x
N_{cell}	Number of cells $[-]$
$\dot{n}$	Molar mass flow rate $[mol/s]$
N_{shaft}	Shaft speed $[rpm]$
P	Power $[W]$
$a_j(\vec{x})$	Polynomial chaos expansion j-th coefficient
p_f	Fin pitch $[m]$

Pr	Prandtl number
p	Pressure $[bar]$
R	Resistance $[\Omega]$
R^2	Normal regression parameter
R^2_{adj}	Adjusted regression parameter
R^2_{pred}	Predicted regression parameter
Re	Reynolds number
RR	Recirculation Ratio $[-]$
SS_B	Sum of squares between
SS_W	Sum of squares within
$STCR$	Steam to carbon ratio
T	Temperature $[K]$
t	Thickness $[m]$
V	Voltage $[V]$
Vol	Volume $[m^3]$
W	Section resistante modulus $[m^3]$

Greek Letters

α	Level of significance
β_A	Area density $[m^2/m^3]$
CFN	Annual cash flow [€]
χ	*Covariance*
e	Emission factor
ϵ_{REC}	Recuperator effectiveness
η	Efficiency $[-]$
I	Incoming [€/unit of time]
$L\dot{H}V$	Primary energy consumption
μ	Mean
ν	*Variance*
ψ	Reference cost [€]
ρ	Density $[kg/m^3]$
σ	Standard deviation

τ_d Design mechanical stress $[Pa]$

$\psi_j(\vec{\xi})$ Polynomial chaos random basis function of j-th node

Subscripts

0 Reference

air air

alt Alternator

an Anode

aux Auxiliary

avg average

cat Cathode

C, fuel Fuel Compressor

cond Condenser

dis Discretized element

ej ejector

elec Electrolyte

eva Evaporator

fuel Fuel

gl global

heat heat

in Inlet

inv Inverter

ohm Ohmic

out Outlet

plant Plant

ps Primary surface

reg regenerator

SH Super heater

side Side

sofc Solid oxide fuel cell

t Turbine

th thermal

Chapter 1

Introduction

1.1 Overall objective

This chapter aims to introduce the reader to the research conducted throughout the course of this study, giving an outline of the motivation that has guided the author along the years and the impact that the resulting work has in the research and praticta fields. At first, the motivation for the study of uncertainty quantification (UQ) in energy systems analysis and design is given. The following section gives an overview on uncertainty, with a focus on the different types of uncertainties that can be found in engineering field. Next, the specific objectives of this Ph.D. study, the research methodology and the contribution of the current work to the literature are presented. Finally, an overview of the chapters in this dissertation is presented.

1.2 Motivation

When considering the real world it becomes immediately clear that most of the parameters that characterize the subject of our study, whatever it would be, are not constant, but are affected by so-called uncertainties. More specifically, a design or manufacturing process or simply the operating conditions of a system have many parameters that are not constant in real world, hence uncertainties are general ubiquitous in analysis and design of highly complex engineering systems. Traditional deterministic design ignores uncertainty or assumes uncertainty as a nominal value and uses safety factors to simplify the design process. This design method may lead to a system that has indeed ideal performance at the design points which is considered during the design process. However, under the real operation conditions, the system can have unexpected performance or significant variations in the performance due to neglecting uncertainties. To deal with these uncertain variables, statistical approaches may be useful to summarize the behaviour of the response and to optimize these random variables.
These approaches may increase the computational expense of the design process significantly if compared to the computational work of deterministic optimization, especially when high fidelity analysis tools are used to improve accuracy. Therefore, it is important to develop and implement computationally efficient robust design methodologies to approach the probabilistic problem in a proper way. The science of

quantitative characterization and reduction of uncertainties in applications is called uncertainty quantification(UQ). The aim of uncertainty quantification is to evaluate the effect that uncertain stochastic inputs, defined by the user, has on the monitored output, to understand the propagating dominant uncertainties (figure 1.1). The main obstacle of this type of study is that the model that represents the problem is a highly non-linear and complex system, in general. In addition, the model, where the UQ is applied, can be expensive in terms of computational time, so it is necessary to have a robust tool that can perform the assesment with as few evaluations as possible and in a non-intrusive approach. The main objective of this approach is to obtain the probabilistic parameters of the response functions without making any modifications to the deterministic code. This approach treats the deterministic code (model) as a "black-box" and approximates the results with formulas based on deterministic code evaluations. It tries to determine how likely certain outcomes are if some aspects of the system are not exactly known.

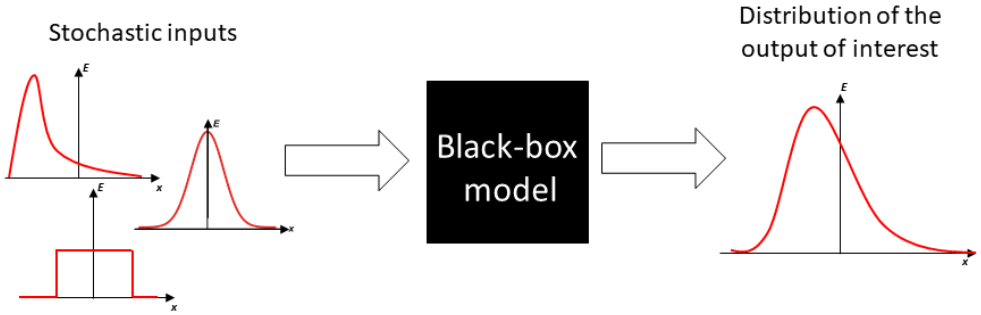


FIGURE 1.1: Scheme of uncertainty propagation

Although closely related, uncertainty quantification and sensitivity analysis are two different disciplines. In fact, we can say that UQ evaluate the uncertainty of the model outputs based on the uncertainties of the input, while sensitivity analysis assesses the contributions of the inputs to the total uncertainty in analysis outcomes. Ideally, uncertainty and sensitivity analysis should be run in tandem in order to get a full understanding of the analysed system.
Uncertainty analysis takes a set of randomly chosen input values (which can include parameter values), passes them through a model (or transfer function) to obtain the distributions (or statistical measures of the distributions) of the resulting outputs. The output distributions can be used to:

- describe the range of potential outputs of the system at some probability level
- estimate the probability that the output will exceed a specific threshold or performance measure target value
- make general inferences, such as:
  - estimating the statistical moments of the outputs (i.e. mean, standard deviation, etc.)
  - estimating the probability the performance measure will exceed a specific threshold
  - assigning a reliability level to a function of the outputs, for example, the range of function values that is likely to occur with some probability
  - describing the likelihood of different potential outputs of the system
  - estimating the relative influences of input variable uncertainties

Uncertainty analisys can be however extremely expensive to be performed on

a complex model such as one which similate the behaviour of an energy system. For this reason, it can be helpful to create a surrogate model (or meta-model) which could mimic the behaviour of the system by approximating it with a simpler model on which uncertainty analysis can be performed in more feasible computational frameworks. Therefore, the use of meta-models can be very advantageous, and can be applied even when little is known about the problem, although it must be kept in mind that if the design space exploration is poor, and the response variable is particularly irregular, the result of the meta-model-assisted optimization can be far from the truth because of the bad estimation of the model coefficients or the choice of an unsuitable model.
Hence, the purpose of this study is to apply some of the main methods for uncertainty propagation and quantification applicable to highly complex systems such as hybrid energy system and combined cycles.
The first part of this work presents an overview of the mathematical framework used in uncertainty quantification, analyzing the possible sources of uncertainty, their classification and the description of the fundamentals of statistics. The main techniques for uncertainty quantification are then analyzed, with a particular focus on Monte Carlo and Response Sensitivity Analysis (RSA), which have been widely used by the author. In particular, these two methods have been firstly applied for the analysis of a recuperator, in order to evaluate the impact of uncertainties in some of the main operating parameters on their cost and volume. From this analysis, the RSA method maintains a good level of accuracy in the results with an important reduction in terms of computational effort.
The Monte Carlo method was also applied to a solid oxide fuel cell hybrid system , which however presented a higher complexity in terms of modeling, making almost impossible, or at least extremely time-consuming, the application of such method directly to the model. For this reason it was decided to make use of the response surface methodology, to create a polynomial metamodel which would resemble the behaviour of the system, allowing however the application of the Monte Carlo method. This system was analyzed both from a performance and both from an economic point of view, to provide useful and complete information on the impact of uncertainties on the system. The same system was studied under different perspectives, showing the strength of the response surface methodology chosen and of the method in general for the study of uncertainties on such systems.
The final application of the Monte Carlo methodology was performed on a combined cycle, preceded by a statistical analysis of gas and electricity prices to create probability density functions of such variables which would mimic their real behaviour, but considering the uncertainties in a future feasibility study of the system.

1.3 Thesis objective, original contributions and pratical impact

The **overall objective** of this doctoral work is to investigate the impact of uncertainties on advanced energy systems, analyzing their impact on the performance and on the economic feasibility of the system, and to develop a methodology to deal with them properly. This approach will be demonstrated by a complete analysis of a solid oxide fuel cell system. In order to accomplish this goal, the following major tasks are envisioned:

- Analyze and compare the uncertainty quantification methodologies to understand and highlight strength and weakness of each of them.
- Apply two of the main uncertainty quantification methods to the study of a recuperator for micro-gas turbines, to evaluate the impact of the uncertainties on the cost and volume of such component.
- Understand and then develop an integrated approach for applying stochastic modeling techniques and uncertainty analysis methodologies to the complex, highly non-linear models; the excessive computational burdens inherent to traditional stochastic approaches (e.g., Monte Carlo simulation) must be resolved before such an integration is practically feasible.
- Study, analyze and compare the different techniques for response surface methodology to apply uncertainty quantification methods on more complex systems.
- Integrate the stochastically based uncertainty approach with response surface methodology into the analysis of a solid oxide fuel cell hybrid system. In particular, a first analysis was performed to evaluate the impact of the uncertainties on the performance of the system. The second analysis was instead focused on the design of such system when considering technological uncertainties and their impact on the performance and on the revenues, finally performing a robust design when considering the optimal designs obtained from a multi-objective optimization. Then, an off-design analysis was carried out to evaluate the effect of the uncertainties on the system working at different loads, considering also an uncertainty in the market in terms of electrical and fuel prices.
- Lay the foundations for a study under uncertainties of a more advanced energy system, such as the integration of Heat Pump and combined cycle, bearing with new scenarios. A preliminary application of the uncertainty quantification techniques and statistical analysis was performed, to evaluate the impact of market uncertainties on the feasibility study of such system.

The **originality** of what is proposed here lies both in the development, implementation, and validation of a comprehensive probabilistic design and optimization approach for the analysis of advanced energy systems.
The **practical impact** of this doctoral reseach on a large-scale energy systems developments is:

- to advance the state-of-the-art beyond the deterministic system development approach by providing analytical solution with known uncertainties
- to provide a robust methodology for dealing with complex energy system subject to uncertainties over a set of different operating conditions
- to provide an evaluation of the impact of the uncertainties in market conditions on complex energy systems, as the growing share of renewable energy sources is already impacting the market in terms of uncertainties.

Hence, the computational efficiency and accuracy of stochastic approach with or without optimization will be demonstrated on advanced energy systems analysis as well as model problems.

1.4 Dissertation outline

The dissertation is composed of seven chapters. In the second chapter the theory that lies behind the uncertainty quantification is presented, as well as a literature review. The focus will be posed, first, on the state of art of current uncertainty quantification technique applied mainly to the energy systems and turbomachinery fields.
Chapter 3 will describe with more details some of the meain uncertainty quantification methods, with a particular focus with the ones used in this thesis work. Sampling methods like Monte Carlo simulation and approximated methods like Response Sensitivity Analysis and Polynomial Chaos will be explained, as well as ANOVA. In the same chapter also the main design of experiment and response surface methodology techniques will be deepened.
Then, in Chapter 4, the application of the Monte Carlo simulation and Response Sensitivity Analysis to an engineering case study will be shown. In particular, these techniques will be applied to compact plate-fin recuperators for micro-turbines applications to analyze the impact of uncertainties in design parameters on their cost and volume.
Next, Section 5 will represent the core of the dissertation, with the application of a methodology for a comprehensive analysis of solid oxide fuel cell hybrid systems subject to technological and market uncertainties. In this section the response surface methodology will be implemented to create a metamodel on which performing a Monte Carlo simulation to estimate the effect of uncertainties both on the performance and both on the revenues of the studied system. In this framework a multi-objective optimization will be performed, and then the optimal solution will be subject to a robust design analysis in order to evaluate the most robust design.
In Section 6 the thermo-economic assessment under market uncertainties of a combined cycle will be treated. The importance of considering the uncertainties in the market in terms of electricity and gas prices is fundamental when analysing the future feasibility of such systems. In this section the author will show a method to re-create an hourly profile of the cost of electricity analyzing historical data and considering the uncertainties in the future market, to evaluate the profitability of the system and the emissions, which represent a critical aspect.
Finally, in the last section all the relevant conclusions and a discussion on future works will be given.

Chapter 2

Uncertainty Quantification

2.1 Introduction

Uncertainty Quantification (UQ) is a very active research area devoted to the study of uncertainties that affect prediction capabilities and can be defined as the science of quantitative characterization and reduction of uncertainties in applications. Quantification of variability and uncertainty is playing an increasing role in regulatory decisions. Although the importance of uncertainty analysis in decisions has increased, it is often the case that such analyses require significant resources of time, money, and expertise. One of the main challenges in this field is the efficiency in propagating uncertainties from the sources to the quantities of interest, especially when there are many sources of uncertainties.
In this work, we will start from an important assumption about our view of the world: we are unable to observe and/or predict nature accurately. In practice, this is due to several causes: the limited accuracy of measurement instruments, the exceedingly high cost of performing accurate experiments and/or measurements, the lack of accurate models or the computational need of using simplified ones and sometimes also the limited knowledge of the phenomena itself. In the field of engineering, predictions are the result of a combination of measurements, modeling and simulations. Analyses are regularly performed based on these predictions and the standard assumption made is that reality will not move significantly away from them. In general non-linear systems show complex behaviours when perturbed and this can lead to unforeseen effects, with the consequence of compromising the particular outcome of an engineering analysis. In engineering, the objectives of such analysis are often related to the minimization of costs, the maximization of revenue, the improvement of safety factors, etc., and unforeseen behaviours can have a dramatic impact on them. In the context of this work, a perturbation is not to be merely intended as an external impulse to the system, but rather as a lack of knowledge on some property of the system. This interpretation of perturbed systems is the reason behind the adoption of probability theory for the description of the possible outcomes of engineering analysis.
Uncertainty Quantification is the mathematical field devoted to the description of uncertainties. Probabilities will help us describing these uncertainties, thus enhancing the capabilities of engineering analysis, enabling the expression of the likelihood of an event to happen. Considering uncertainties on a system, we have actually ignored a non-negligible assumption: the mathematical model used to describe the natural system is correct. In spite of having physical evidences about the outcome

of systems, many of the commonly adopted models are based on certain levels of approximation, introduced to make them better manageable. A model is a mathematical entity describing the causality connection between some input and some output. The models we are going to investigate here are steady-state or dynamical systems, i.e. time-space dependent models, where the inputs are usually boundary conditions and/or initial values, while the outputs are the states of the system. These models, as well as their inputs, may very well depend on parameters p which affects their outputs. In order to have any predictive capability, models need to be evaluated, associating particular input conditions to particular output states. Many models have closed form solutions which enable a fast analytic evaluation, but the majority of them need to be simulated by complex, time consuming and ultimately expensive computational techniques. Despite the advances in numerical methods for these simulations and the overgrowing computing power, the attained incredible accuracy may very well be spoiled by the comparison with reality: the input conditions, parameters and output states are often known within "engineering accuracy", known as probability distributions or in some cases not even known. Even if the model uncertainty is a critical problem for many engineering applications, in this work we will not address it directly.

Several theories address the definition of uncertainty. These theories include probability theory [1, 2], fusy set theory [3] and evidence theory [4, 5]. In this thesis, we work under the framework of probability theory, which provides a solid and comprehensive theoretical foundation and offers the most versatile statistical tools. In contrast to the traditional, deterministic simulations, we describe uncertainties as randomness, and model the sources of uncertainties as random variables, random processes and random fields. To quantify the sources of uncertainties, we must specify the joint probability density function of all these random variables, processes and fields. This step is usually very problem-dependent. The methods involved in this step include statistical analysis, experimental error analysis and often expert judgment [6]. Although how to quantify model uncertainties and numerical uncertainties is still a topic of current research [7, 8] , successful examples exist of quantifying the uncertainty sources for very complex engineering systems.

Once the sources of uncertainties are quantified, we need to calculate how these uncertainties propagate through the simulation to the quantities of interest. These, also known as objective functions, are the main quantities to be predicted. They are functions of all the random variables that describe the sources of uncertainty. The final product of the uncertainty qualification process is a quantitative description of the likelihood in the values of the quantities of interest. This can only be obtained by combining our knowledge of the sources of uncertainties and the behaviour of the objective functions with respect to these sources. In the probability theoretic framework, this quantitative description is a joint probability density function of the objective functions. The support of this joint probability density function, i.e., the space where the function is positive, describes all possible scenarios predicted by the computational simulation; in addition, the value of the probability density function indicates how likely each scenario is. This joint probability density function enables decision making based on risk analysis, removing the important limitations of deterministic computational simulations.

In the following the sources of uncertainties are briefly described as well as come of the main methods for uncertainty analysis. The methods which have been used by the author during his work instead will be analyzed deeper in the following chapters.

2.2 Quantification of uncertainties

Uncertainties arise in all aspects of mechanical design. In order to identify and quantify uncertainties, we first need to distinguish between errors and uncertainties. The uncertainty quantification community has introduced precise definitions to characterize various types of uncertainties. The following definition can be found in [9], which provides guidelines for the verification and validation of CFD simulations.
Uncertainty is defined as: *"A potential deficiency in any phase or activity of the modeling process that is due to the lack of knowledge"*.
Error is defined as: *"A recognizable deficiency in any phase or activity of modeling and simulation that is not due to lack of knowledge"*.
The key words in the definition of uncertainty are potential deficiency and lack of knowledge. Since the sources of the uncertainties are mostly unknown, the characterization of uncertainties is a difficult task. Hence, this definition of uncertainties suggests a stochastic approach. On the other hand, the identification of errors is a more straightforward task due to the knowledge of the error sources.

2.2.1 Errors vs uncertainties

The American Institute of Aeronautics and Astronautics (AIAA) Guide for the Verification and Validation of CFD Simulations defines errors as recognizable deficiencies of the models or the algorithms employed and uncertainties as a potential deficiency that is due to lack of knowledge [9]. This definition is not completely satisfactory because does not precisely distinguish between the mathematics and the physics. It is more useful to define errors as associated to the translation of a mathematical formulation into a numerical algorithm (and a computational code). Errors are typically also further classified in two categories: acknowledged errors are known to be present but their effect on the results is deemed negligible. On the other end, unacknowledged errors are not recognizable but might be present; implementation mistakes (bugs) or usage errors can only be characterized by comprehensive verification tests and procedures. Using the present definition of errors, the uncertainties are naturally associated to the choice of the physical models and to the specification of the input parameters required for performing the analysis. As an example, numerical simulations require the precise specification of boundary conditions and typically only limited information is available from corresponding experiments and observations. Therefore, variability, vagueness, ambiguity and confusion are all factors that introduce uncertainties in the simulations.
Another important distinction to highlight is the one between variability and uncertainty. In fact, variability and uncertainty are two distinct concepts within a decision problem, even though they often have been lumped together in environmental analyses. Variability results from natural stochastic behaviour, periodicity, or variance in a trait across a population. In contrast, uncertainty is related to lack of knowledge about the "true" value of a quantity, lack of knowledge regarding which of several alternative models best describes a mechanism of interest, or lack of knowledge about which of several alternative probability density functions should represent a quantity of interest [10]. If variability and uncertainty are unaccounted for, the quality of environmental assessment will be potentially affected. Similarly, failure to account for uncertainty may lead to assumptions of precision that do not convey

the true state of knowledge. Thus, both variability and uncertainty may have ramifications on policy assessments and the ultimate success of the resulting policies. Understanding which sources of variability and uncertainty are reducible provides insight to the process of determining how to most appropriately allocate resources to improve the certainty in the results of an analysis. Understanding the sources of variability and uncertainty is hence critical in identifying their existence and characterizing their impact.

2.2.2 Sources of variability

Variability is present in any dataset that is not static in space, time, or across members of a population. Common sources of variability are stochasticity, periodicity, population variance, and existence of subpopulations. These sources are briefly discussed based on [10].
Stochasticity is random, non-predictable behaviour that is common in many physical phenomena. Stochasticity is irreducible, although it can typically be represented over time or space with a frequency distribution. Use of averaging periods can also reduce the effect of stochasticity.
Periodicity implies cyclical behaviour. For example, ambient temperature is cyclical, tending to rise during the day and decrease at night. Periodicity is sometimes addressed using averaging periods or assessing maximum values. Time series approaches can also be used to represent periodicity more explicitly [11].

2.2.3 Sources of uncertainty

It is often stated that uncertainty is a property of the analyst [10]. Different analysts, with different states of knowledge or access to different datasets or measurement techniques, may have different levels of uncertainty regarding the predictions that they make. Sources of uncertainty can include problem and scenario specification, model uncertainty, random error, systematic error, lack of representativeness, lack of empirical basis, and disagreement of experts. The first two sources of uncertainty are typically related to structural uncertainty, while the rest are related to uncertainty in model inputs.

- **Problem and Scenario Specifications**. Typically, a scenario includes specification of goals and scope of an environmental assessment such as assessment boundaries. However, in some cases a scenario may fail to consider all of the factors and conditions contributing to variation in the output, and hence, uncertainty can be introduced. This source of uncertainty typically results in a bias in estimates. Important factors may be omitted from assessment because of lack of available resources.
- **Model Uncertainty**. Typically, the computer models used in assessment do not capture all aspects of the problem. There is a trade-off between model formulation and scope of the assessment. Some sources of uncertainty in modeling

include conceptual model, model structure, mathematical implementation, detail, resolution, and boundary conditions). The analyst's conceptual model of the problem may be different from reality. For example, the analyst may inadvertently omit important inputs, emphasize some inputs, or misrepresents dependence between some inputs. There may also be competing model formulations that may have advantages or disadvantages over each other. The alternative formulations may also represent different sets of thought or assumptions. Model uncertainty due to mathematical implementation of the model may result from numerical instability. There also a possibility that coding errors introduces additional uncertainty. Model detail refers to the degree to which physical and chemical processes are described within the model. Typically, the assumption is that through having a model with greater level of detail, less uncertainty should appear in results. However, in some cases more detailed models may require the introduction of new parameters, the values of which may be uncertain. Thus, increasing detail may not reduce overall uncertainty or may lead to greater uncertainty. Model resolution refers to factors such as the time steps and grid sizes used. Decreasing these parameters can yield more detailed results. Boundary conditions may be based on expert judgment, assumed behaviour, or the outputs from other models. Each source can potentially contribute to uncertainty. Simulation models generally become less sensitive to initial conditions as the simulation time frame is increased, but they can retain an important sensitivity to boundary conditions at all-time steps.

- **Random Error/Stochasticity**. It represents random, non-predictable behaviour. This source of uncertainty is associated with imperfections in measurement techniques or with processes that are random or statistically independent of each other [12]. For example, in a time series analysis, the behaviour of a time-dependent trait is characterized with a function. However, this function does not capture the whole variability. Typically, an error term with a normal distribution is used to represent the unexplained (random) behaviour.

- **Systematic Errors**. The mean value of a measured quantity may not converge to the "true" mean value because of biases in measurements and procedures. Systematic error can be introduced into a dataset from sources such as the imperfect calibration of equipment, simplifying or incorrect assumptions, or errors due to the selection and implementation of methodologies for collecting and utilizing data. Application of surrogate data is another source of systematic error [12]. This occurs when one assumes a simplified model for how a system behaves and then makes measurements accordingly.

- **Lack of Representativeness**. Data used as input to the assessment may not be completely representative of the study objectives. Surrogate data is an example of non-representative data. When data describing an input to an assessment is unavailable, limited, or not practical to collect, surrogate data is often used.

- **Lack of Empirical Basis**. This type of uncertainty cannot be treated statistically, because it requires prediction about something that has yet to be built, or measured. This type of uncertainty can be represented using technically-based judgments about the range and likelihood of possible outcomes [10]. These judgments may be based on theoretical foundation or experience with analogous systems. In cases where data exist for analogous systems, it may be possible to fit probability distribution models to the existing dataset.

- **Disagreement of Experts**. Expert opinion is used to select values or distributions for inputs into an assessment. For example, experts may suggest the most appropriate reaction rate or supply a subjective prior distribution for an input. However, different opinions of experts may differ on these data and distributions. When there are limited data or alternative theoretical bases for modeling a system, experts may disagree on interpretation of data or on their estimates regarding the range and likelihood of outcomes for empirical quantities. This disagreement introduces uncertainty regarding the most appropriate values of distributions to use [10].

Uncertainty can enter mathematical models and experimental measurements in various contexts. Another way to categorize the sources of uncertainty that can be considered is:

- **Parameter uncertainty**, which comes from the model parameters that are inputs to the computer model (mathematical model) but whose exact values are unknown to experimentalists and cannot be controlled in physical experiments, or whose values, cannot be exactly inferred by statistical methods.
- **Parametric variability**, which comes from the variability of input variables of the model. For example, the dimensions of a work piece in a process of manufacture may not be exactly as designed and instructed, which would cause variability in its performance.
- **Structural uncertainty**, or model inadequacy, model bias, or model discrepancy, which comes from the lack of knowledge of the underlying true physics. It depends on how accurately a mathematical model describes the true system for a real-life situation, considering the fact that models are almost always only approximations to reality. One example is when modeling the process of a falling object using the free-fall model; the model itself is inaccurate since there always exists air friction. In this case, even if there is no unknown parameter in the model, a discrepancy is still expected between the model and true physics.
- **Algorithmic uncertainty**, or numerical uncertainty, which comes from numerical errors and numerical approximations per implementation of the computer model. Most models are too complicated to solve exactly. For example the finite element method or finite difference method may be used to approximate the solution of a partial differential equation, which, however, introduces numerical errors. Other examples are numerical integration and infinite sum truncation that are necessary approximations in numerical implementation.
- **Experimental uncertainty**, or observation error, which comes from the variability of experimental measurements. The experimental uncertainty is inevitable and can be noticed by repeating a measurement for many times using exactly the same settings for all inputs/variables.
- **Interpolation uncertainty**, which comes from a lack of available data collected from computer model simulations and/or experimental measurements. For other input settings that don't have simulation data or experimental measurements, one must interpolate or extrapolate in order to predict the corresponding responses.

Errors and approximations in input data measurement, parameter values, model structure and model solution algorithms are all sources of uncertainty. While there are reasonable ways of quantifying and reducing these errors and the resulting range of uncertainty of various system performance indicator values, they are impossible to eliminate. Decisions will still have to be made aware of a risky and uncertain future, and can be modified as new data and knowledge are obtained in a process of adaptive management.

2.3 Classification of uncertainties

Differences between model outputs and observed values can result from either natural variability, and/or by both known and unknown errors in the input data, the model parameters or the model itself. The latter is sometimes called knowledge uncertainty, but it is not always due to a lack of knowledge. Models are always simplifications of reality and, hence, "imprecision" can result. Sometimes imprecision occurs because of a lack of knowledge, such as just how a particular species will react to various environmental and other habitat conditions. At other times, known errors are introduced simply for practical reasons. Imperfect representation of processes in a model constitutes model structural uncertainty. Imperfect knowledge of the values of parameters associated with these processes constitutes model parameter uncertainty. Natural variability includes both temporal variability and spatial variability, to which model input values may be subject. In order to identify and characterize the uncertainties in the energy systems framework, we now introduce the following classification of uncertainties, which is generally accepted in literature [13, 14, 15]: inherent (aleatory) uncertainty and model-form (epistemic) uncertainty. The **aleatory uncertainties** are the uncertainties, which cannot be reduced because we do not have control on them. The **epistemic uncertainties** are uncertainties, which could be reduced by more accurate measurements and better modeling of the physics. This distinction is not sharp because it depends on the problem setting and on subjective judgment. From the practical point of view, the distinction of which uncertainty is reducible and which is not, depends on the problem setting. The most relevant example is laboratory experiments, where the explicit aim is to reduce the number of aleatory uncertainties to a minimum. Thus, some uncertainties, which are considered aleatory in a field experiment, would be considered as epistemic in the corresponding laboratory experiment.

2.3.1 Aleatory uncertainty

Aleatory uncertainty arises because of natural, unpredictable variation of parameter values, initial and boundary conditions, geometry, etc. in the performance of the system under study. Depending upon the application, there may be numerous sources of aleatory uncertainty within a physical system. The knowledge of experts cannot be expected to reduce this kind of uncertainty although their knowledge may be useful in quantifying the uncertainty. The determination of material properties or operating conditions of a physical system typically leads to aleatory uncertainties; additional experimental characterization might provide more conclusive evidence

of the variability but cannot eliminate it completely.
Thus, this type of uncertainty is sometimes referred to as irreducible uncertainty. The determination of material properties or operating conditions of a physical system typically leads to aleatory uncertainties; additional experimental characterization might provide more conclusive evidence of the variability but cannot eliminate it completely. Aleatory uncertainty is normally characterized using probabilistic approaches. Since this uncertainty is due to the random nature of input data and can be mathematically characterized by a Probability Density Function (PDF) if there is enough information on the type of the distribution.
Selecting the most appropriate and accurate distribution types for random input variables is important because it can have a significant impact when propagating the input uncertainty to the uncertainty in the output variable of interest. Hence, implicit in any uncertainty analysis are the assumptions that statistical distributions for the input values are correct and that the model is a sufficiently realistic description of the processes taking place in the system. Neither of these assumptions is likely to be entirely correct.
This type of uncertainty was also categorized by Phadke [16] into three groups:

- External: External sources of performance variation, independent of the product itself (ex. Environment)
- Unit-to-unit variation: variation due to manufacturing variability (or manufacturing tolerances)
- Deterioration: Variation due to the change in product performance with time and wear

The perception of this type of uncertainty can be quantified using conventional statistical methods. The source of aleatory uncertainty can be distinguished from other forms by depiction as randomly distributed quantities.

2.3.2 Epistemic uncertainty

Epistemic uncertainty is what is indicated in the AIAA guide [9] as uncertainty, hence a potential deficit that is due to a lack of knowledge. The key feature of this definition is that the primary cause is incomplete information of some characteristics of the system. It can arise from assumptions introduced in the derivation of the mathematical model used or simplifications related to the correlation or dependence between physical processes. Examples are the lack of experimental data to characterize material or poor understanding of coupled physics phenomena. These uncertainties are sometimes characterized as “state of knowledge” uncertainties, which mean that the uncertainties can be reduced with more knowledge of the physical process.
Hence, epistemic uncertainty is due to incomplete information or ignorance about the behaviour of the system that is conceptually resolvable. It is obviously possible to reduce the epistemic uncertainty by using, for example, a combination of calibration, inference from experimental observations, expert judgment and improvement of the physical models. As a result, epistemic uncertainty is described as reducible uncertainty. The use of mathematical models and the choice of model assumptions

can contribute to this effect. The resolution from the numerical grid cell size of a model offers an approximation to system behaviour, which can be a source of epistemic uncertainty. The simplification of complex nonlinear systems using linear approximate models gives rise to this type of uncertainty [17].
For the characterization of epistemic uncertainty, the statistical distribution types are not suitable to be used to describe the nature of the epistemic parameter due to the lack of knowledge or information about the uncertainty. Studies conducted by Oberkampf [13] and Helton [2] show that the modeling of epistemic uncertainty with probabilistic approaches may lead to inaccurate predictions for uncertainty in the responses, due to the lack of information on the characterization of uncertainty as probabilistic. One approach to model the epistemic uncertainty is to characterize it with intervals. In this approach, the upper and lower bounds on the uncertain variable can be prescribed using either limited experimental data or expert judgment. All values within this interval are equally likely to occur because it is not appropriate to assign a statistical distribution to epistemic uncertain variables. Examples of epistemic uncertainties associated with aerodynamic simulations can include the value of turbulence modeling parameters and fluid transport quantities. Hence, epistemic uncertainty is not well characterized by probabilistic approaches because it might be difficult to infer any statistical information due to the nominal lack of knowledge.

2.4 Probability distribution estimation

The uncertainty on the parameters driving the model output needs to be quantified accurately in order for the UQ analysis to be reliable. The probability distribution estimation of the parameters can be carried out in several ways:

- by assumption: probability distributions are assigned to the parameters relying on experience and wisdom, by measurement: extensive measurements of the parameters are carried out and probability distributions are fitted to these experiments
- by inference: the probability distributions are reconstructed using measurements of the quantity of interest (QoIs).

The quality of the UQ analysis carried out will depend strongly on the quality of the distributions constructed. Once probability distributions are available for all the input quantities in the computational algorithm, the objective is to compute the PDFs of the output quantities of interest. This step is usually the most complex and computationally intensive for realistic engineering simulations. A variety of methods are available in the literature, from sampling based approaches (e.g. Monte Carlo) to approximated methods (e.g. Response Sensitivity Analysis or Polynomial Chaos). In the next chapters, some of these methods will be described in detail, with a particular focus on the ones which have been used by the author.

2.5 Propagation of uncertainties

There are two major types of problems in uncertainty quantification: **forward uncertainty propagation**, which is the most frequently used, and **backward uncertainty propagation** [18]. The work carried out by the author will only focus on the forward uncertainty propagation, which is the quantification of uncertainties in system outputs propagated from stochastic inputs.

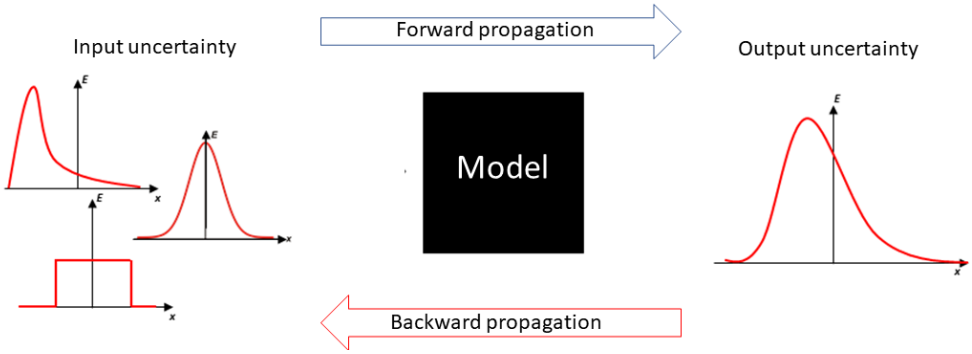


FIGURE 2.1: Schematic description of the different types of propagation of uncertainties

The targets of uncertainty propagation analysis can be:

- to evaluate low-order moments of the outputs, i.e. mean and variance,
- to evaluate the reliability of the outputs. This is especially useful in reliability engineering where outputs of a system are usually closely related to the performance of the system,
- to assess the complete probability distribution of the outputs. This is useful in the scenario of utility optimization where the complete distribution is used to calculate the utility.

2.6 Description of statistics

Statistics is a mathematical science pertaining to the collection, analysis, interpolation or explanation, and presentation of data [19]. Statistical methods can be used to describe a collection of data and this is called descriptive statistics. In addition, patterns in the data may be modelled in a way that accounts for randomness and uncertainty in the observations, and then are used to draw inferences about the process or population being studied; this is called inferential statistics. Descriptive, predictive, and inferential statistics comprise applied statistics [20]. In the uncertainty quantification framework, the statistics is important to predict the behaviour of the system. This is because that the concept of robust design is to minimize the variation in the performance of the system as well as to pull the mean performance onto the target value. If input parameters are characterized using PDFs, the mean and variance of input parameters can be derived based on their types of PDFs. The

mean is the centre of area of the distribution – summarizing the location properties of the distribution. The variance is the second moment of area of the distribution about the mean - summarizing the way in which the area is spread over the object. The r^{th} moment about the origin, also called raw moment, of a random variable X, denoted by $\mu_r^{'}$ is the expected value of X^r, simbolically expressed as:

$$\mu_r^{'} = E\left(X^r\right) = \sum_x x^r f(x) \tag{2.1}$$

for $r = 0, 1, \ldots n$ when X is discrete, and

$$\mu_r^{'} = E\left(X^r\right) = \int_{-\infty}^{\infty} x^r f(x)\, dx \tag{2.2}$$

when X is continuous. For r=1, the raw moment is $\mu_1^{'} = E\left(X\right) = \mu_x$, which is the mean of the distribution of X, or simply the mean of X.
When the r^{th} moment is not about the origin, but is about the mean of a random variable X, then is denoted by μ_r and is the expected value of $(X - \mu_X)^r$, which can be written also as:

$$\mu_r = E\left[(X - \mu_x)^r\right] = \sum_x (x - \mu_X)^r f(x) \tag{2.3}$$

for $r = 0, 1, \ldots n$ when X is discrete, and

$$\mu_r = E\left[(X - \mu_x)^r\right] = \int_{-\infty}^{\infty} (x - \mu_X)^r f(x)\, dx \tag{2.4}$$

when X is continuous. The r^{th} moment about the mean is defined if $E\left[(X - \mu_x)^r\right]$ exists. The second moment about the mean is called **variance** and is mathematically expressed as:

$$\mu_2 = E\left[(X - \mu_x)^2\right] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)\, dx = Var(X) = \sigma^2 \tag{2.5}$$

Its unit is the square of the unit of the random variable. Hence, it always has a positive value. It measures the spread of the distribution. Thus, a zero variance implies a deterministic variable.
The positive square root of the variance is the **standard deviation**, which is also considered a "second moment" quantity and is usually indicated with σ. The ratio between the standard deviation and the mean is called Coefficient Of Variance (COV), and is a measure of the relative spread of the distribution. This ratio is dimensionless and so is often used to cast formulae in a dimesionless form.

2.7 Uncertainty quantification techniques

Engineering design commonly assumes nominal values for uncertain parameters to simplify the design process. Unfortunately, this simplification can lead to a product that exhibits performance worse than nominal in real-world conditions. The accurate solution of physical problems in terms of mathematical models requires a fully understanding of all the basic phenomena involved and their detailed description, in general, in terms of ordinary or partial differential equations. Such models are formulated in terms of physical properties or constants that, in general, contain some level of uncertainties. Therefore, to obtain more realistic models and results that more closely reflect the real world, uncertainty should be considered more formally (as opposite to through sensitivity analysis) in the performance/design and operation/control simulation of energy systems. In order to quantify this uncertainty, the distribution of model inputs should be propagated through the model to obtain distributions on model outputs. The simulation of advanced energy systems has widely used for understanding the performance of such systems. However, it is difficult to analyse this kind of system taking into consideration their variability because traditional probabilistic approaches are very computationally expensive even though such probabilistic approaches for uncertainty analysis can significantly improve the quality of the analysis results. Traditional probabilistic approaches include Monte Carlo simulation (MC) which is a typical sampling method, which requires a large volume of samples in order to obtain the probabilistic information (i.e. the mean, variation, skewness, and probability distribution function). This also required a large amount of computer memory. Such a large volume of samples makes it prohibitive to apply sampling methods on advanced energy systems. To get around this problem, a number of approximate probabilistic modeling approaches have been proposed in the literature, which can be grouped into several types of approaches: Response Sensitivity Analysis, Polynomial chaos, Mid-range Approximation Methods and Response Surface. These methods provide approximate solutions that can closely approximate the MC result, which is able to provide relatively speaking the exact solution. The advantage of these methods is that they guarantee highly effective computational time for which large-scale energy system is crucial. Some of these approaches are explained in detail in Chapter 3 and a state of art is briefly described below in the following section.
In general, uncertainty propagation techniques may be analytical, approximated, or numerical.

2.7.1 Analytical Propagation Methods

For simple models in which the output is a function of linear combination of model inputs with no dependency, the propagation of probability distributions through the model is straightforward. For such cases, the Central Limit Theorem (CLT) can be used. CLT states that if we have a population with mean μ and standard deviation σ and take sufficiently large random samples from the population with replacement, then the distribution of the sample means will be approximately normally distributed. To take a sample "with replacement" means that when a unit is selected at random from the population it is returned to the population, and then another element is selected at random [21].

- **Advantages**: Analytical propagation techniques based on the CLT are straightforward and easy to implement to simple models in which the model output is a linear sum or products of model inputs
- **Disadvantages**: Although the results of the CLT approach are useful in some cases for propagating the mean and the variance through a simple linear model, they do not imply anything about the shape of the model output distribution. Moreover, the implications of the CLT are relevant only if the conditions of the CLT exist for a particular situation. Thus, if a model contains both products and sums of inputs, or for which some of the inputs are dominant over others, or for which some of the inputs are not statistically independent, the analytical propagation techniques based on the CLT cannot be used.

2.7.2 Approximation Methods based on Taylor Series

There are a number of methods based upon the use of the Taylor series expansions for propagating the mean and other central moments of random variables through a model. The basic approach is to take a general function, such as:

$$y = h(x_1, x_2, \ldots, x_n) \tag{2.6}$$

and then expand the function about the point $[E(x_1), E(x_2), \ldots, E(x_n)]$ using a multivariate Taylor series expansion. The series is usually truncated at a specified set of higher order terms. For example, the mean of the output can be approximated as [22]:

$$E(y) = h\left[E(x_1), E(x_2), \ldots, E(x_n)\right] + \frac{1}{2}\sum_{i=1}^{n}\frac{\partial^2 \overline{h}}{\partial x_i^2}\sigma_{x,i}^2 + [high\ order\ terms] \tag{2.7}$$

The function h and the partial derivatives on the right side are evaluated at the point $[E(x_1), E(x_2), \ldots, E(x_n)]$. The variance of the model output, σ_y^2, of statistically independent random variables is approximated by the following:

$$\sigma_y^2 = \sum_{i=1}^{n}\left(\frac{\partial \overline{h}}{\partial x_i}\right)^2 \sigma_{x,1}^2 + \sum_{i=1}^{n}\left(\frac{\partial \overline{h}}{\partial x_i}\right)^2 \frac{\partial^2 \overline{h}}{\partial x_i^2}\mu_3\left(x_i\right) + [high\ order\ terms] \tag{2.8}$$

where $\mu_3\left(x_i\right)$ is the third central moment of each input random variable.

- **Advantages**: based upon a sufficient number of central moments for a model output, it may be possible to select a parametric probability distribution model that provides a good representation of the output distribution [22]. Once a

parametric distribution of the output is specified, prediction can be made regarding any percentile of the model output. Thus, as an advantage of approximation methods based upon Taylor series, it may only be necessary to propagate the moments of each probability distribution of the model inputs instead of the entire probability distributions

- **Disadvantages**: approximation methods based on Taylor series typically have three major limitations [22]. First, as a primary limitation in application of these techniques, the model function should be differentiable. Therefore, these methods cannot be applied to problems with discrete or discontinuous behaviours. Second,these methods are computationally intensive as they typically require the evaluation of second order (and potentially higher) derivatives of the model. Third, although these techniques are capable of propagating central moment of input distributions, information regarding the tails of the input distributions cannot be propagated.

2.7.3 Numerical Propagation Techniques

The most common techniques for numerical propagation of uncertainty and variability are sampling based methods. Some of the sampling based methods for propagating probability distributions are Monte Carlo, Fourier Amplitude Sensitivity Test (FAST) and Reliability Based Methods. Among those Monte Carlo is by far the most well-known and broadly used. In Monte Carlo simulation, a model is run repeatedly, using different values for each of the uncertain input parameters each time [23, 22, 24]. The values of each of the uncertain inputs are generated based on the probability distribution for the input. With many input variables, one can envision Monte Carlo simulation as providing a random sampling from a space of m dimensions, where m is the number of inputs to a model. As a general approach for applying Monte Carlo simulation to a model, for each input a probability distribution should be specified. Random samples are generated from the each of the probability distributions. One sample from each input distribution is selected, and the set of samples is fed into the model. The model is then executed as it would be for any deterministic analysis. The process is repeated until the specified number of model iterations has been completed. Thus, instead of obtaining a single number for model outputs as in a deterministic simulation, a set of samples is obtained. These can be represented as CDFs and summarized using typical statistics such as mean and variance. Most numerical simulation methods, including random Monte Carlo, require the generation of uniformly distributed random numbers between 0 and 1 [22]. Given a uniformly distributed random variable, several methods exist from which to simulate random variables that are described by other probability distributions (e.g., normal, lognormal, and gamma). These methods include the inverse transform, composition, and function of random variables [24]. In addition, methods exist for simulation of jointly distributed random variables, which enables one to represent correlations between two or more simulated random variables.

- **Advantages**: their output provides more information compared to analytical and approximate methods. Moreover, because Monte Carlo methods provide a probability distribution of the output, they avoid the problem of compounding conservative values of input variables. Additional advantages follow from

the information provided by Monte Carlo simulation. For example, results based on Monte Carlo simulations are typically conducive to sensitivity analysis, permitting the risk assessors to determine where additional data will be most useful in reducing uncertainty [25, 26]

- **Disadvantages**: because Monte Carlo simulation requires multiple iterations of a model, such simulations can be computationally intensive if the model requires a large run time per simulation. Furthermore, depending on the data quality objectives of the analysis, it may be necessary to perform a large number of simulations. Although not a limitation of the method itself, in practice results based on Monte Carlo simulations are easy to misuse by stretching them beyond the limits of credibility. For example, problems can arise when inexperienced analysts use commercial simulation packages due to ease of application and lack of familiarity with underlying assumptions and restrictions. Typical misapplications of Monte Carlo simulation include failure to properly develop input distributions and misinterpretation or over-interpretation of results. For example, it is not possible to have a precise estimate of an upper percentile of an output distribution without a large simulation sample size.

2.8 State of the art

An extensive literature on this argument has appeared in the last 70 years, since Stanislaw Ulman came up with the Monte Carlo method in 1946 [27]. As shown in Table 2.1, many methods have appeared during the years, addressing different issues. The list includes some of the methods that will be presented in Chapter 3, without the ambition to present all the methods available today. As it is often the case in numerical methods, the main issues encountered in the propagation of uncertainties are related to the achievement of a good balance between accuracy and time consumed. Since *"there ain't no such thing as a free lunch"*, all the methods presented have strengths and weaknesses, thus methods applied to particular problems need to be accurately selected.

TABLE 2.1: Methods for propagation of uncertainty and approximate year of appearance

Name	Year	Reference
Wiener chaos expansion	1938	[28]
Monte Carlo method	1946	[27]
Quasi-Monte Carlo method	1961	[29]
Latin Hypercube	1979	[30]
Response Sensitivity Analysis	1981	[31]
Multi-Point Approximation Methods	1996	[32]
Sparse Grid Quadrature	1998	[33]
Generalized Polynomial Chaos	2002	[34]

A comparative analysis of the main methodologies for uncertainty propagation and quantification has been carried out by Cuneo, Traverso and Shahpar [23], identifying the strength and weakness of such methods when applied to the engineering

field. Only a limited number of applications of any of these methods have been applied in the literature to the performance/design and operation of energy systems. For example, Subramanyan and Diwekar [35, 36], Suo and Huang [37] and Hart and Jacobson [38] have applied the MC approach to some aspect of uncertainty related to the design and/or operation of energy systems, while Gorla [39] has done the same but using another approximated method called Fast Probability Integration (FPI). The author himself has made use of the Monte Carlo method to study of the propagation of uncertainties in a recuperator [40] and in hybrid systems [41, 42], which will be discussed in more details in the following chapters.
Related to the Polynomial Chaos, Montomoli et al. [43] have studied the uncertainties in film cooling geometry due to manufacturing processes. Panizza et al. [44] applied the sparsegrid approach on the uncertainty quantification of a centrifugal compressor performance and demonstrated the effectiveness of it compared to Monte Carlo. Abraham et al. [45, 46] presented a non-intrusive regression-based method for building sparse PC expansions, demonstrating its strength for CFD applications considering geometrical and operational uncertainties, while Ghisu et al. [47] evaluated the variability in the performance of a generic modular-core compression system for a three-spool modern gas turbine engine subject to uncertain operating conditions. Polynomial Chaos Expansion was also applied by Coppitters et al. [48] coupled with a multi-objective optimization algorithm to a photovoltaic-eloctrolyzed system subject to techno-economic uncertainties.
The Multi-Point Approximation Method was used by Shahpar and Caloni [49] to carry out a high-fidelity design optimization, and adapted by Korolev et al. [50] to solve large scale engineering optimization problems with uncertainty in the design variables and in additional variables which cannot be addressed by the designer, showing its potential against some benchmark optimization problems such as the ten-bar truss and the cantilever beam.

References

[1] L. L. Green, H. Z. Lin, and M. R. Khalessi. "Probabilistic methods for uncertainty propagation applied to aircraft design". In: *20th AIAA Applied Aerodynamics Conference*. 2002. ISBN: 9781624101106. DOI: 10.2514/6.2002-3140.

[2] J. C. Helton and J. D. Johnson. "Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty". In: *Reliability Engineering and System Safety* 96.9 (2011), pp. 1034–1052. ISSN: 09518320. DOI: 10.1016/j.ress.2011.02.013.

[3] H.-J. Zimmermann. *Fuzzy Set Theory—and Its Applications*. 2001. DOI: 10.1007/978-94-010-0646-0.

[4] H. R. Bae, R. V. Grandhi, and R. A. Canfield. "An approximation approach for uncertainty quantification using evidence theory". In: *Reliability Engineering and System Safety* (2004). ISSN: 09518320. DOI: 10.1016/j.ress.2004.01.011.

[5] H. R. Bae, R. V. Grandhi, and R. A. Canfield. "Sensitivity analysis of structural response uncertainty propagation using evidence theory". In: *Structural and Multidisciplinary Optimization* 31.4 (2006), pp. 270–279. ISSN: 1615147X. DOI: 10.1007/s00158-006-0606-9.

[6] S. F. Wojtkiewicz, M. S. Eldred, R. V. Field, A. Urbina, and J. R. Red-Horse. "Uncertainty quantification in large computational engineering models". In: *19th AIAA Applied Aerodynamics Conference*. 2001. DOI: 10.2514/6.2001-1455.

[7] C. J. Freitas. "The issue of numerical uncertainty". In: *Applied Mathematical Modelling* 26.2 (2002), pp. 237–248. ISSN: 0307904X. DOI: 10.1016/S0307-904X(01)00058-0.

[8] D. Draper. "Assessment and Propagation of Model Uncertainty". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (1995), pp. 45–70. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1995.tb02015.x.

[9] AIAA. "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations". In: *AIAA Journal* (1998). ISSN: 03064379. DOI: 10.2514/4.472855.001.

[10] A. Cullen and H. Frey. *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. 1999. DOI: 10.1002/sim.958.

[11] L. Todd. *The Oxford Handbook of Quantitative Methods, Volume 2 Statistical analysis*. 2013. ISBN: 9788578110796. arXiv: arXiv:1011.1669v3.

[12] P. R. Bevington, D. K. Robinson, J. M. Blair, A. J. Mallinckrodt, and S. McKay. "Data Reduction and Error Analysis for the Physical Sciences". In: *Computers in Physics* 7.4 (1993), p. 415. ISSN: 08941866. DOI: 10.1063/1.4823194.

[13] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. "Error and uncertainty in modeling and simulation". In: *Reliability Engineering and System Safety* 75.3 (2002), pp. 333–357. ISSN: 09518320. DOI: 10.1016/S0951-8320(01)00120-X.

[14] A. D. Kiureghian and O. Ditlevsen. "Aleatory or epistemic? Does it matter?" In: *Structural Safety* 31.2 (2009), pp. 105–112. ISSN: 01674730. DOI: 10.1016/j.strusafe.2008.06.020.

[15] O. De Weck, C. Eckert, and J. Clarkson. "A classification of uncertainty for early product and system design". In: *Proceedings of ICED 2007, the 16th International Conference on Engineering Design*. Vol. DS 42. 2007. ISBN: 1904670024.

[16] M. Phadke. *Quality Engineering Using Robust Design*. 1995, p. 250. DOI: 10.2307/1269049.

[17] R. L. Winkler. "Uncertainty in probabilistic risk assessment". In: *Reliability Engineering and System Safety* 54.2-3 (1996), pp. 127–132. ISSN: 09518320. DOI: 10.1016/S0951-8320(96)00070-1.

[18] R. Ghanem, H. Owhadi, and D. Higdon. *Handbook of uncertainty quantification*. Ed. by Springer. Berlin, Germany, 2017, pp. 1–2053. ISBN: 9783319123851. DOI: 10.1007/978-3-319-12385-1.

[19] Alain Desrosieres. *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, Massachussetts, and London, England, 1998. ISBN: 0674689321.

[20] B. J. Winer. *Statistical principles in experimental design*. 2009. DOI: 10.1037/11774-000.

[21] S. G. Kwak and J. H. Kim. "Central limit theorem: The cornerstone of modern statistics". In: *Korean Journal of Anesthesiology* 70.2 (2017), pp. 144–156. ISSN: 20057563. DOI: 10.4097/kjae.2017.70.2.144.

[22] G. J. Hahn and S. S. Shapiro. *Statistical Models in Engineering*. Wiley Classic Library, 1994.

[23] A. Cuneo, A. Traverso, and S. Shahpar. "Comparative analysis of methodologies for uncertainty propagation and quantification". In: *Proceedings of the ASME Turbo Expo*. Vol. 2C-2017. American Society of Mechanical Engineers, 2017. ISBN: 9780791850800. DOI: 10.1115/GT2017-63238. URL: https://asmedigitalcollection.asme.org/GT/proceedings/GT2017/50800/Charlotte,NorthCarolina,USA/241782.

[24] M. Morgan, M. H. Granger, and M. Small. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Vol. 8. 1. 1990. ISBN: 0521427444. DOI: 10.1016/0169-2070(92)90021-z.

[25] J. C. Helton. "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal". In: *Reliability Engineering and System Safety* 42.2-3 (1993), pp. 327–367. ISSN: 09518320. DOI: 10.1016/0951-8320(93)90097-I.

[26] I. M. Sobol. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation* 55.1-3 (2001), pp. 271–280. ISSN: 03784754. DOI: 10.1016/S0378-4754(00)00270-6.

[27] N. Metropolis and S. Ulam. "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247 (1949), pp. 335–341. ISSN: 1537274X. DOI: 10.1080/01621459.1949.10483310.

[28] N. Wiener. "The Homogeneous Chaos". In: *American Journal of Mathematics* 60.4 (1938), p. 897. ISSN: 00029327. DOI: 10.2307/2371268.

[29] H. Niederreiter. "Quasi-Monte Carlo methods and pseudo-random numbers". In: *Bulletin of the American Mathematical Society* 84.6 (1978), pp. 957–1042. ISSN: 0002-9904. DOI: 10.1090/s0002-9904-1978-14532-7.

[30] M. D. McKay, R. J. Beckman, and W. J. Conover. "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 15372723. DOI: 10.1080/00401706.1979.10489755.

[31] M. D. Dettinger and J. L. Wilson. "First order analysis of uncertainty in numerical models of groundwater flow part: 1. Mathematical development". In: *Water Resources Research* 17.1 (1981), pp. 149–161. ISSN: 19447973. DOI: 10.1029/WR017i001p00149.

[32] V. V. Toropov and V. L. Markine. "The use of simplified numerical models as mid-range approximations". In: *6th Symposium on Multidisciplinary Analysis and Optimization*. 1996, pp. 952–958. DOI: 10.2514/6.1996-4088.

[33] T. Gerstner and M. Griebel. "Numerical integration using sparse grids". In: *Numerical algorithms* 18.3 (1998), pp. 209–232. ISSN: 1017-1398. DOI: 10.1023/A:1019129717644.

[34] D. Xiu and G. Em Karniadakis. "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos". In: *Computer Methods in Applied Mechanics and Engineering* 191.43 (2002), pp. 4927–4948. ISSN: 00457825. DOI: 10.1016/S0045-7825(02)00421-8.

[35] K. Subramanyan and U. M. Diwekar. "Characterization and quantification of uncertainty in solid oxide fuel cell hybrid power plants". In: *Journal of Power Sources* 142.1-2 (2005), pp. 103–116. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2004.09.030.

[36] U. M. Diwekar, E. S. Rubin, and H. C. Frey. "Optimal design of advanced power systems under uncertainty". In: *Energy Conversion and Management* 38.15-17 (1997), pp. 1725–1735. ISSN: 01968904. DOI: 10.1016/s0196-8904(96)00184-7.

[37] M. Q. Suo, Y. P. Li, and G. H. Huang. "Multicriteria decision making under uncertainty: An advanced ordered weighted averaging operator for planning electric power systems". In: *Engineering Applications of Artificial Intelligence* 25.1 (2012), pp. 72–81. ISSN: 09521976. DOI: 10.1016/j.engappai.2011.08.007.

[38] E. K. Hart and M. Z. Jacobson. "A Monte Carlo approach to generator portfolio planning and carbon emissions assessments of systems with large penetrations of variable renewables". In: *Renewable Energy* (2011). ISSN: 09601481. DOI: 10.1016/j.renene.2011.01.015.

[39] R. S. Gorla. "Probabilistic analysis of a solid-oxide fuel-cell based hybrid gas-turbine system". In: *Applied Energy* 78.1 (2004), pp. 63–74. ISSN: 03062619. DOI: 10.1016/S0306-2619(03)00006-0.

[40] A. Giugno, A. Cuneo, and A. Traverso. "Analysis of uncertainties in compact plate-fin recuperators for microturbines". In: *Applied Thermal Engineering* 150.September 2018 (2019), pp. 1243–1251. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2019.01.093. URL: https://doi.org/10.1016/j.applthermaleng.2019.01.093.

[41] A. Cuneo, L. Mantelli, A. Giugno, and A. Traverso. "Uncertainty Quantification Analysis of a Pressurised Fuel Cell Hybrid System". In: *Proceedings of ASME Turbo Expo 2019: Turbomachinery Technical Conference and Exposition*. Vol. 91351. Phoenix, 2019, pp. 1–15.

[42] A. Giugno, L. Mantelli, A. Cuneo, and A. Traverso. "Performance analysis of a fuel cell hybrid system subject to technological uncertainties". In: *Applied Energy* 279 (2020). ISSN: 03062619. DOI: 10.1016/j.apenergy.2020.115785.

[43] F. Montomoli, M. Massini, S. Salvadori, and F. Martelli. "Geometrical uncertainty and film cooling: Fillet radii". In: *Journal of Turbomachinery* 134.1 (2011). ISSN: 0889504X. DOI: 10.1115/1.4003287.

[44] A. Panizza, D. T. Rubino, and L. Tapinassi. "Efficient uncertainty quantification of centrifugal compressor performance using polynomial chaos". In: *Proceedings of the ASME Turbo Expo*. Vol. 2B. 2014. ISBN: 9780791845615. DOI: 10.1115/GT2014-25081.

[45] S. Abraham, M. Raisee, G. Ghorbaniasl, F. Contino, and C. Lacor. "A robust and efficient stepwise regression method for building sparse polynomial chaos expansions". In: *Journal of Computational Physics* 332 (2017), pp. 461–474. ISSN: 10902716. DOI: 10.1016/j.jcp.2016.12.015.

[46] S. Abraham, P. Tsirikoglou, C. Lacor, G. Ghorbanias, D. Wunsch, C. Hirsch, and F. Contino. "Uncertainty quantification in industrial turbo-machinery design using sparse polynomial chaos expansions". In: *2018 Multidisciplinary Analysis and Optimization Conference*. 2018. ISBN: 9781624105500. DOI: 10.2514/6.2018-3103.

[47] T. Ghisu, G. T. Parks, J. P. Jarrett, and P. J. Clarkson. "Adaptive polynomial chaos for gas turbine compression systems performance analysis". In: *AIAA Journal* (2010). ISSN: 00011452. DOI: 10.2514/1.J050012.

[48] D. Coppitters, W. De Paepe, and F. Contino. "Surrogate-assisted robust design optimization and global sensitivity analysis of a directly coupled photovoltaic-electrolyzer system under techno-economic uncertainty". In: *Applied Energy* 248 (2019), pp. 310–320. ISSN: 03062619. DOI: 10.1016/j.apenergy.2019.04.101.

[49] S. Shahpar and S. Caloni. "Aerodynamic optimisation of high pressure turbines for lean-burn combustion system". In: *Proceedings of the ASME Turbo Expo*. 2012. ISBN: 9780791844748. DOI: 10.1115/GT2012-69228.

[50] Y. M. Korolev, V. V. Toropov, and S. Shahpar. "Design optimization under uncertainty using the multipoint approximation method". In: *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2017*. 2017. ISBN: 9781624104534. DOI: 10.2514/6.2017-1934.

Chapter 3

Stochastic Methods for the analysis of Energy Systems

3.1 Introduction

All models of real systems exhibit such uncertainties to one extent or another. Therefore, to obtain more realistic models and results that more closely reflect the real world, uncertainty should be considered more formally (as opposite to through sensitivity analysis) in the performance, design and operation/control of energy systems.
Uncertainty quantification has been widely used for decision-making procedures in work done on product management and scheduling. However, it is difficult to apply to large-scale system because traditional probabilistic approaches are very computationally expensive. Even though such probabilistic approaches for uncertainty analysis can significantly improve the quality of the analysis results, only a few cases have been reported in the literature applied to energy system.
In engineering design under uncertainty, the uncertainties are usually modelled using probability theory. In Reliability-Based Design Optimization (RBDO), variations are represented by standard deviations, which are typically assumed constant, and a mean performance is optimized subject to probabilistic constraints. In general, probability theory is very effective when sufficient data is available to quantify uncertainty using probability distributions. However, when sufficient data is not available or there is lack of information due to ignorance, the classical probability methodology may not be appropriate. For example, during the early stages of product development, quantification of the product's reliability or compliance to performance targets is practically very difficult due to insufficient data for modelling the uncertainties. A similar problem exists when the reliability of a complex system is assessed in the presence of incomplete information on the variability of certain design variables, parameters, operating conditions, boundary conditions etc.
Probabilistic and Randomized Methods for Design under Uncertainty examines uncertain systems in control engineering and general decision or optimisation problems for which data is not known exactly. Uncertainty in model parameters depends on the model granularity level. For instance, model parameters in transient simulation models describe the physical behaviour of an individual component while those in normative models describe the characteristics of systems at an aggregate

level. Accordingly, uncertainty associated with different levels of model parameters should be separately investigated. Hence, quantification uncertainty is accomplished by the three steps:

1. investigating physics-based equations that parameterize the behaviour of aggregate-level parameters with a set of detailed model parameters;
2. quantifying uncertainty in the detailed model parameters from the literature review or experimental data;
3. propagating quantified uncertainty through selected equations to derive a probability distribution for one aggregate-level parameter.

As discussed in Section 2, Uncertainties in the results of system can be affected by several uncertainty resources, which can be categorized into direct and indirect sources. Since not incorporating these uncertainties into the modelling process may produce misleading results, it is important to consider uncertainty effects on system simulation, design and operation/control.

3.2 Probabilistic design methods

A deterministic model (i.e. one without uncertainty considerations) or nondeterministic/probabilistic model (i.e. one with uncertainty considerations) can be described in general terms by:

$$\vec{M} = f\left(\vec{X}, \vec{Y}\right) \tag{3.1}$$

where the vector $\vec{M}$ represents a set of system output values, and vectors $\vec{X}$ and $\vec{Y}$ correspond to a set of design and operation/control variables, respectively. If the model is a non-deterministic or probabilistic one and the uncertainties on and are known, their effects on system development and performance can be evaluated via several possible probabilistic design methods, which can be divided into sampling method and approximate methods, i.e.:

- **Sampling methods**
  - Monte Carlo (MC) simulation
  - Adaptive Sampling
  - Importance Sampling
  - etc.
- **Approximate methods**
  - Response Sensitivity Analysis (RSA)
  - Polynomial Chaos (PC)
  - Fast Probability Integration (FPI)
  - etc.

Traditional probabilistic approaches include Monte Carlo simulation (MC), which is a typical sampling method requiring a large number of repeated simulations in order to obtain the probabilistic information (i.e. the mean, variation, skewness, and probability distribution function). However, when there is a large number of degrees of freedom are being used to determine the optimal operation/control of the system, the MC is so computationally intensive that combined with large-scale optimization it renders the problem computationally unfeasible. Because of this weakness, other modified sampling methods have been proposed; but these are still inadequately modified for being able to apply them to large-scale system simulation and optimization.
This computational difficulty can be overcome by the use of approximated methods like Response Sensitivity Analysis (RSA), Polynomial Chaos (PC) and Response Surface (RS). These methods provide approximate solutions that can closely approximate the MC result, which is able to provide probably the exact solution. The advantage is that they guarantee highly effective computational time for which large-scale energy system optimization is crucial. In fact, these are approximate methods, which can only approach the results extracted from sampling methods such as MC, which provides exact solutions as the number of samples approaches infinity. Of course, how close the approach is and how robust the convergence to a solution determines the viability of the approximate methods.
It is important to underline that all these methods require that the input variables are independent and so uncorrelated. Independency means that their joint probability distribution is the product of their marginal probability distribution. On the other hand, uncorrelation means that their correlation coefficients are zero. If the input variables are correlated, they can easily be made uncorrelated, taking into consideration their covariance [51, 52], but the higher is the number of variables the higher is the complexity to make them uncorrelated. For example, to apply PC when the inputs are correlated, some methods have been proposed, i.e. linear transformation [53], L-L expansion [54] or proper orthogonal decomposition [55]. All of these fix the problem by applying transformation to remove the correlations, which increase the complexity of the problem and degrades the convergence rate because of increased non-linearity [56]. Response Sensitivity Analysis can take into account the correlation through the variables, considering the covariance between them in the evaluation of the output variance [57]. However, it is common to assume that the uncorrelated random variables resulting from the transformation can be treated as independent. This assumption is valid for uncorrelated standard normal variables, but may be an approximation for uncorrelated standard uniform, exponential, beta, and gamma variables. Hence, in this way, the UQ methods could be applied. On the other hand, if the correlation between the input variables is not taken into account in a proper way, a large error in the results can occur.

3.2.1 Monte Carlo

The Monte Carlo algorithms family is widely used within numerical mathematics, statistics, and UQ. Historically, the idea of this algorithm appeared in the 1940s, having John von Neumann, Stanislas Ulam, and Nicolas Metropolis as its inventors [27]. In a real world system, design and operation/control variables show characteristics of randomness with a certain degree of uncertainty. This uncertainty can be described by random numbers corresponding to particular probabilistic distributions.

MC can then be used (and has widely been used) to determine the propagation of this uncertainty in the computational analysis of various physical and mathematical problems. MC is particularly distinguished as a probabilistic simulation technique since it can solve extremely complex and discontinuous problems precisely, provided the model is simulated with a high enough sampling number. Once the probabilistic information of a variable such as the mean value, the variance, and/or the probability distribution is known, the randomness of the variable can be simulated close to its true or real randomness using a random number generator. Such a generator is based on a computational algorithm and, thus, is only a pseudo-random number generator. This is because computational algorithms produce long sequences of apparently random results, which are in fact completely determined by a shorter initial value, known as a seed value or key. Repeating the simulation of the analyzed systems based on a set of randomly generated input variable values and storing the system output values, a set of probabilistic values and probability distribution functions (PDFs) of the output variables are obtained. This type of generator is useful for simulating the real randomness of the variables in computational studies.
In MC n random samples are generated based on a prescribed probability density function (PDF), where n is user-defined and depends on the problem itself. Afterwards, each generated sample is the input in the numerical solver of the system and the resulting numerical equation is solved. Note that after the stochastic input is plugged into the model of the system, the problem becomes deterministic. Finally, after all samples are propagated, the last step consists of post-processing, such as the statistical evaluation of the PDF for each output (i.e. mean, variance, standard deviation, etc.).

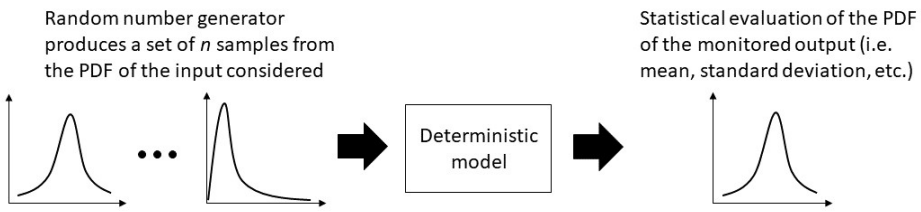


FIGURE 3.1: Schematization of Monte Carlo approach

This type of approach explicitly results in exact uncertainty (relatively speaking) propagation from the input variables to the system responses, assuming that the sampling number is high enough. Moreover, the probabilistic information of the system outputs quantify the range in the confidence level with which the performance of the system can be viewed relative to the objective limits set on the system. However, even though MC produces exact solutions and is a powerful probabilistic design method for complex nonlinear energy systems, it is not very practical because of the very large computational effort or burden required. In fact, even if this method is really robust and versatile, it has a slow convergence rate with a probabilistic error of $o(1/\sqrt{M})$, where M represents the number of samples. It features a major disadvantages, this is its accuracy depends on large number of executions, making it computationally expensive. However, it is important to remark that the convergence of MC is weakly dependent of the number of dimensions [58], thus, when the number of dimension is high (typically >50) it is a very good choice. When only few numbers of samples are available for the analysis, a useful approach to estimate the confidence intervals of the moments is the Bootstrap method [59]. When only limited samples are available, the bootstrap method can provide an efficient way of

estimating the distribution of statistical parameter using the re-sampling technique [60]. The idea is to create many sets of bootstrap samples by re-sampling with replacement from the original data. The size of the initial samples is n and the number of bootstrap resamplings is p. Each re-sampling can be performed by sampling with replacement n data out of the n initial samples (hence, the bootstrap samples contain repeated values from the initial samples and omit some of the initial values). Since the re-sampling process draws samples from the existing set of samples, it does not require additional simulations. Since the re-sampling procedure allows for selecting data with replacement, the statistical properties of the re-sampled data are different from those of the original data. This approach allows estimating the distribution of any statistical parameter without requiring additional data. The standard error or confidence intervals of the statistical parameter can also be estimated from the bootstrap distribution. However, the bootstrap method provides only an approximation of the true distribution because it depends on the values of the initial samples.
To overcome the computational problem of traditional MC, several improvements of this method have been developed, which can be classified into the following two classes: variance reduction techniques and quasi-random or low-discrepancy sampling [61, 58]. Other techniques are related to new efficient Design of Experiment (DoE) sampling techniques, e.g. Latin Hypercube [30], Halton or Hammersley samplings [62, 63], have been used. A second option for reducing the computational cost, used in optimization procedures, is to combine MC methods with a surrogate or approximate model, which allows a cheaper evaluation of the deterministic output (e.g. Kriging methods [64]). An alternative sampling method, on which the focused was posed in the last few years, is the Quasi-Monte Carlo (QMC) method. Recently quasi-Monte Carlo algorithms have been successfully used for very large values of dimensions, especially in financial applications; see [65, 66]. The errors for these examples were observed to be independent of d and were of order $n-1$. Hence, quasi-Monte Carlo algorithms win in two ways over Monte Carlo, in that we have both a better exponent of convergence and a better assurance of error. On the other hand, the quasi Monte Carlo methods seek to construct nodes that perform significantly better than the average [67], which means that the nodes are deterministically chosen, such that a small error is guaranteed. The error bound of the method can be improved to $O\left(M-1\left(\log M\right)^{d-1}\right)$, if the nodes are properly chosen [67]. The main difference between Monte Carlo methods and quasi Monte Carlo methods is that quasi Monte Carlo methods are completely deterministic, thus the error bounds are also deterministic. Moreover, the quasi Monte Carlo method can significantly reduce the computational effort. Detailed information about these methods can be found in [67, 66, 68]. QMC methods are known to work better than MC methods when the integrand is sufficiently smooth, whereas they can completely fail on an integrand of unbounded variation [69]. Quasi-Monte Carlo method uses low discrepancy sequences in order to uniformly covers the sampling domain. Unlike pseudo-random sequences, quasi-random sequences do not attempt to imitate the behaviour of random sequences. Instead, the elements of a quasi-random sequence are correlated to make them more uniform than random sequences, and hence faster convergence for quadrature formulas. This can provide convergence rates of $O\left(\left(\log M\right)^{d}/M\right)$ where d is the dimension of the random space. Some have objected to the name quasirandom, since these sequences are intentionally not random. Because of the correlations, quasi-random sequences are less versatile than random or pseudo-random sequences. They are designed for integration, rather

than simulation or optimization. On the other hand, the desired result of a simulation can often be written as an expectation, which is an integral, so that quasi-Monte Carlo is then applicable.
Another promising new approach is based on a Multi-Level Monte Carlo (MLMC) method e.g. as introduced by Giles [70] and it shows great potential for CFD related stochastic analysis. This method is a powerful tool in the context of uncertainty quantification for hyperbolic systems of conservation laws. The method appears to be flexible and can be used for different types of uncertain inputs such as random initial data, source terms or flux functions. Moreover, the MLMC method can deal with a very large number of sources of uncertainty. For instance, a computation for shallow water equations with uncertain bottom topography is reported, involving approximately 1000 sources of uncertainty, while being several orders of magnitude more efficient than the standard Monte Carlo method, without applying surrogate models [71]. The MLMC method is based on the use of different levels of accuracy for the deterministic analysis and a different number of sampling points for each one. The statistical representation of the result of the MLMC method is ensured by a high number of sampling points analysed with a low level of accuracy whereas the discretization error is controlled by a short number of sampling points analysed with a high level of accuracy. This combination provides very good quality results for the stochastic analysis with a much more reasonable computational cost compared with classical MC method.
In conclusion, even though MC produces exact solutions and is a powerful probabilistic design method for complex nonlinear energy systems, it is not very practical because of the very large computational effort required. This is because in engineering application, one function evaluation is very expensive. The sampling methods, even the improved methods, are not an appropriate choice in our case to compute statistics of the flow, since they do not offer the possibility to improve the approximation quality by additional knowledge of the underlying function, e.g. smoothness information.

Mean Square Pure Error Methodology

In simulation models using the Monte Carlo method, the Experimental Error hence varies with the variation of the sample's size and depends on the number of times the simulation is replicated, namely the function of the number of replicated runs. Many researchers studying these issues recommend a rather large number of replicated runs. In conventional DoE and RSM applications, the experimental campaign following the design phase is carried out on the real system from which some samples are usually taken to be analyzed in terms of objective function or, in some cases, by measuring the responses directly in the system itself. The Sum Squares Error (SSE) whose amount comprises all the experiment-related and model-related errors expressed in terms of Sum Squares for Pure Error (MLMC) and Lack-of-Fit Sum Squares (SSLOF) respectively, allows to determine the fraction of the Sum Squares Total (SST) or the total variability that still escapes the experimenter's comprehension [72]. Figure 3.2 shows how the error is spread between the experimental phase and the system modelling phase.

As stated above, SS_{PE} is the pure error component and, as such, it cannot be controlled by the experimenter. It is expressed by:

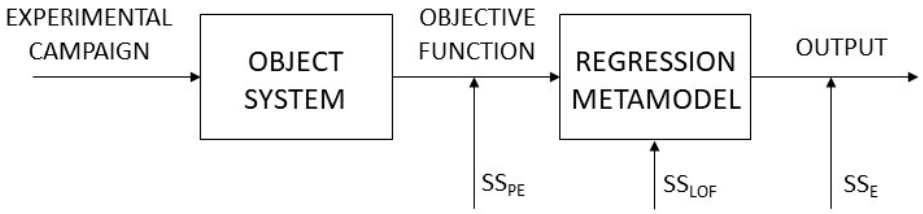


FIGURE 3.2: Schematic decomposition of the experimental error

$$SS_{PE} = \sum_{t=1}^{a} \sum_{u=1}^{n_t} (y_{t,u} - \overline{y}_t)^2 \tag{3.2}$$

$$MS_{PE} = \frac{SS_{PE}}{N - a} \tag{3.3}$$

where u stands for the experimental responses replicated at the same experimental level, y_i is the mean of the observations under the same i-th level of the independent variables, N is the total number of observations, a is the number of the levels of the independent variables, and n_i is the number of replications under the i-th level. SSLOF is the error component due to the model's lack of fit and namely the component ascribable to regression:

$$SS_{LOF} = \sum_{t=1}^{a} (n_t (\hat{y}_t - \overline{y}_t))^2 \tag{3.4}$$

$$MS_{LOF} = \frac{SS_{LOF}}{a - (p + 1)} \tag{3.5}$$

being $p + 1$ the total number of β regressors appearing in the meta-model.
When the system is, instead, a complex industrial plant working by its own very nature in a discrete mode and in the presence of high stochasticity, for obvious reasons of convenience, costs and time, the intervention carried out directly on it to test new operating hypotheses is avoided and it is hence preferred to build a new simulation model in which new experimental tests are to be done. This way it is possible to avoid any interference with the operating of the real system and to test also especially innovative and, at times, even daring hypotheses, obtaining faster responses with almost negligible costs thanks to the computation power of electronic processors. While the advantages of this approach are absolutely clear, the need to translate the real system into a simulation model generates, however, with regard to what has been said in the previous paragraph, a double component of additional error that must be taken into account (see figure 3.3):

- the first component is directly related to the transcription of reality into the simulation model
- the second is linked instead to the transformation of the pure error that changes

from invariant time into developmental in the simulated time (or in the replicated runs) as per the formula below [73, 74]:

$$SS_{PE}(t_i) = \sum_{j=1}^{n_0} \left(y_j(t_t) - \overline{y}(t_i)\right)^2 \tag{3.6}$$

$$MS_{PE}(t_i) = \frac{SS_{PE}(t_i)}{n_0 - 1} \tag{3.7}$$

where n_0 is the project's number of experimental replications, namely of replicated runs of the simulation model. According to the theory, pure error can be calculated at the center of the experimental domain by replicating the so-called central tests [46].

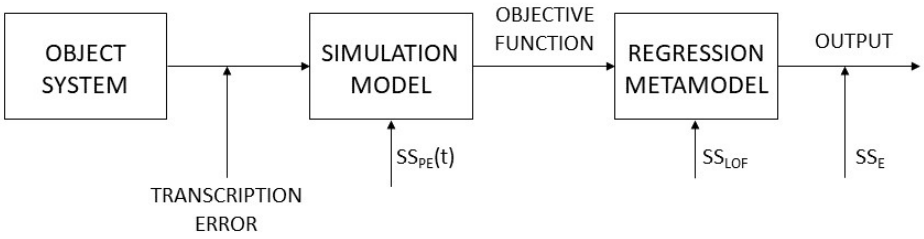


FIGURE 3.3: Schematic decomposition of the experimental error in simulation applications

Under these conditions, Mean Square Pure Error (MSPE) takes on, in the simulated time, the shape of knee-shaped curve whose trend tends asymptotically towards the background noise in the real system. Choosing a run time or a number of experimental replications capable of ensuring the smallest possible MSPE compatibly with the level of stochasticity typical of the real system, with the time available for the delivery of the results and the Lack of Fit becomes then crucial in MCs. This important phase becomes decisive when facing optimization problems and it determines the outcome of the project chosen, as it affects the MSLOF and MSPE [74].
According to the universally accepted definition, a Monte Carlo simulator is a teller of the possible histories of the object system. The "goodness" model depends not only on its constructor ability (i.e., system analysis, data survey, and logic transcription) but also on a correct experimental activity, which should include, among its main targets, experimental error measurement, which is generally distributed as a normal distribution $(0, \sigma^2)$, affecting the model [75, 72, 76]. The σ^2 entity, which can be estimated according to Cochran's theorem [72] through the measurement of the mean square pure error (MSPE), its unbiased estimator, is an intrinsic characteristic of each model. σ^2 is strictly connected to the investigated reality, since it is directly dependent on the overall stochasticity by which this reality is affected. The study of the MSPE trend makes it possible to solve the problem through a graph whose examination clearly points out, without any particular difficulty in interpretation, the total model noise fraction in each simulated run [77].
In order to obtain the evaluation of MSPE results, it is necessary to study the evolution in the replicated runs both of the variance of the sampling distribution of the sample mean $(MSPE_{MED})$ and the variance of the sampling distribution of the sample standard deviation $(MSPE_{STDEV})$. These two parameters, taken together, allow choosing the numbers of runs needed to obtain an unbiased evaluation of the experimental error affecting the objective function. As it is well known, the larger are the

samples; the better is the description of the population. With this methodology, it is possible to graphically highlight the evolution of the variance experimental error as a function of the sample size.
For the experimenter, the problem does not lie in obtaining a theoretical $MS_{PE} = 0$, which by result of the central limit theorem can be obtained for a sample of infinite amplitude, but rather in limiting the number N of runs by a thorough check of the evolution in the experimental error in terms both of magnitude and stabilization so as to limit also its impact on the response to acceptable values. This way the experimenter will be able to choose the best trade-off between the experimental cost and the expected results. Furthermore, the knowledge in each point of the values of $MSPE_{MED}$ and $MSPESTDEV$ allows carrying out important inferences on the behaviour of the real experimental response, which can vary between a minimum value and a maximum value.

TABLE 3.1: Collection of experimental data for MSPE calculation

Experimental Table Runs			
$y_{1,1}$	$y_{1,2}$	$\cdots$	$y_{1,K}$
$y_{2,1}$	$y_{2,2}$	$\cdots$	$\cdots$
$y_{3,1}$	$y_{3,2}$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$
$y_{N,1}$	$y_{N,2}$	$\cdots$	$y_{N,K}$

The technique for MSPE study in replicated runs, can be divided into the following phases:

1. set a number $K > 2$ of simulation runs, carried out in parallel, in which the independent model variables are maintained always at the same level, modifying only the triggering seeds of the random numbers. In the case of a single replication factorial experiment or central composite design application, K will be equal to the central runs used in the experimental design (in this regard, it should be borne in mind that the variance of the pure experimental error must be constant in each point of the operability region and hence at the center as well as along the boundary)

2. determine a number $N >> 1$ of replications y_{ij} with $i = 1, \ldots, N$, $j = 1, \ldots, K$ for each simulation run (see Table 3.1)

3. calculate for each of the K runs N means, defined as $\overline{y_{ij}}$ with $i = 1, \ldots, n, \ldots, N$ where:
$$\overline{y_{nj}} = \frac{\sum_{i=1}^{n} y_{ij}}{n} \tag{3.8}$$

4. calculate N means of the means, defined as $\overline{Y_i}$ with $i = 1, \ldots, n, \ldots, N$, given by:
$$\overline{Y_n} = \frac{\sum_{j=1}^{K} \overline{y_{nj}}}{K} \tag{3.9}$$

5. calculate N values of the $MSPE_{MED}$, with $i \leq j \leq K$ and with $1 \leq i \leq N$, as shown in Table 3.2:

$$MSPE_{MED}(i) = \frac{\sum_{j=1}^{k} \left(\overline{y_{ij}} - \overline{Y_i}\right)^2}{K-1} \tag{3.10}$$

These values, transferred on the plane $(i, MSPE_{MED})$, highlight the trend of the mean square pure error curve in the replicated runs, thus showing, step by step, the entity of the error variance affecting the simulation trial by impacting each objective function. As mentioned above, according to Cochran's theorem, $MSPE_{MED}$ is the best estimator of the experimental error variance σ^2 and, consequently, allows for the measurement of the experimental error affecting the mean value of the means distributions.

TABLE 3.2: Means calculation for MSPE

Means Table Runs				Row Mean
$\overline{y_{1,1}} = y_{1,1}$	$\overline{y_{1,2}}$	$\cdots$	$\overline{y_{1,K}}$	$\overline{Y_1} = \frac{\sum_{j=1}^{K} \overline{y_{1,j}}}{K}$
$\overline{y_{2,1}} = \frac{y_{1,1} + y_{2,1}}{2}$	$\overline{y_{2,2}}$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	
$\overline{y_{N,1}} = \frac{\sum_{i=1}^{N} y_{i,1}}{N}$	$\cdots$	$\cdots$	$\overline{y_{N,K}} = \frac{\sum_{i=1}^{N} y_{i,K}}{N}$	$\overline{Y_n} = \frac{\sum_{j=1}^{K} \overline{y_{N,j}}}{K}$

The same approach is also valid for the standard deviation.
The knowledge of the MSPE values in each point makes it possible to obtain important inferences on the behaviour of the real experimental responses. In fact, always by the effect of Cochran's Theorem, it makes it possible to know the interval in which there is a 99.7% probability that the value $y*$ from a single simulation run lies in it:

$$\begin{aligned} \overline{y} - 3\sqrt{MSPE_{MED}} - 3\sqrt{\overline{VAR} + MSPE_{STDEV}} \leq y* \\ \leq \overline{y} + 3\sqrt{MSPE_{MED}} + 3\sqrt{\overline{VAR} + MSPE_{STDEV}} \end{aligned} \tag{3.11}$$

where $\overline{VAR}$ is the square of $\overline{stdev_N}$. Moreover, it should be noted that when each sample of experimental responses, resulting from K parallel runs, would be broad enough to provide an exhaustive description of the population behaviour, the two $MSPEs$ evolving in the runs would crash on the $X-$axis $(MSPE = 0)$. This way the totality of the stochastic description of the real system, and, hence, in the model, is represented by the experimental response variance. If considered:

$$\lim_{N \to \infty} MSPE_{MED} = \lim_{N \to \infty} MSPE_{STDEV} = 0 \tag{3.12}$$

then equation 3.11 becomes:

$$\overline{y}_{MED} - 3\sqrt{\overline{VAR}} \leq y^{*,\infty} \leq \overline{y}_{MED} + 3\sqrt{\overline{VAR}} \tag{3.13}$$

For the experimenter, the problem is not to obtain a theoretical $MSPE = 0$ but to limit the number of runs N through a careful check of the experimental error evolution in terms of both of magnitude and adjustment, so as to limit also its impact on y^* to acceptable values.
As concerns the number of K runs carried out in parallel, it is clear that the interest to choose a high K value can be correct. In fact, the higher the K, the wider the sample used for the operation. Therefore, the K size necessarily affects the accuracy of the mean of the dependent variable mean/variance distribution. In many cases, despite the computational power available, it may happen that, as K grows, the time to calculate MSPE quickly becomes heavy.
It is an obvious consequence that the study of the experimental error evolution, and the resulting search for the characteristic background noise, can be particularly burdensome, due to the lack of a careful setting of the various parameters. While, for the purpose of a correct evaluation of both the stochastic effect on the experimental response and the characteristic noise, it is important for N to choose values having a size of 10^4 or greater.
In summary, the steps necessary for a correct experimental campaign with the Monte Carlo method are:

1. Construction of the two $MSPE_{MED}$ and $MSPE_{STDEV}$ curves as a function of replicated runs as shown in Figure 3.4
2. Identification of the number of runs necessary for both curves to achieve stabilization based on the analysis of the two graphs. It should be noted that, until the curve does not reach an appropriate stabilization phase, the error magnitude may even change from one experimental campaign to another, leading, under the same conditions, to extremely different simulation outputs
3. Use of the number of runs determined above for the model experimentation phase.

3.2.2 Response Sensitivity Analysis

One of the most practical methods of uncertainty analysis is the approximate Response Sensitivity Analysis (RSA), in which the output parameters of the system can be found with the use of series Taylor expansions. This type of approach is necessary because the simulations based on deterministic models may provide incomplete and often erroneous assessments of the results output of the system. For this reason, in order to get information as complete as possible, it is necessary to be able to obtain indications on the sensitivity (variation) and the uncertainty related to the results. It can be said that, in contrast to deterministic analysis, which does not provide any answers on the interval of validity of information or on the level of uncertainty of the results, the RSA method, guarantees the obtaining of these fundamental information

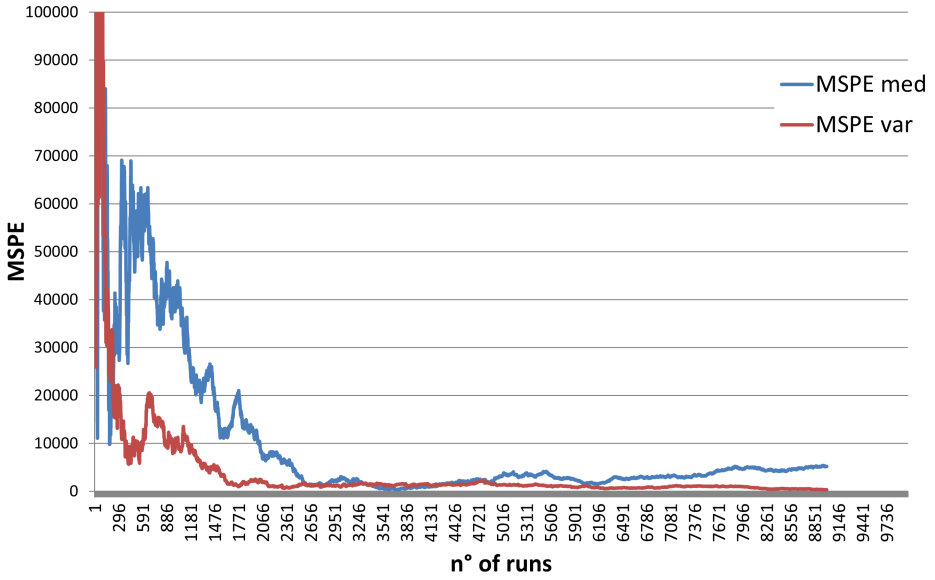


FIGURE 3.4: $MSPE_{MED}$ and $MSPE_{STDEV}$ evolution curves

relying of an analysis system that uses the Taylor series expansions applied to the sensitivity of the system response [57]. The moment of the first order (mean value) and the moment of the second order (standard deviation) of the system outputs can be easily estimated by the development in Taylor series.
If the system has highly nonlinear behaviours and distributions of input values are nonnormal, the RSA method can make mistakes when calculating the results. However, being a method computationally not expensive (in contrast to the MC) it can be easily applied for the analysis of energy systems.
Using the RSA method, there are only two cases in which the probabilistic (PDF information, μ, σ) for system outputs ($\vec{M}$) can be estimated with precision, they are:

- when the explicit relationship ($g_{M_j}(\vec{Z})$) that correlated each of the system's output value (M_j) to the input values ($\vec{Z}$) and also if every probability distribution ($f(Z_i)$), be it normal or non-normal, corresponding to each single input ($\vec{Z}$) is statistically independent of each other is known, it is possible to determine numerically or analytically probabilistic information for each system output. For the j-th parameter (M_j) the PDF will be given by:

$$f_{M_j} = f\left(M_j\right) = f\left(\vec{Z}\right) = f\left(Z_1\right)\left(Z_2\right)\ldots\left(Z_n\right) \tag{3.14}$$

while the mean value by:

$$\mu_{M_j} = \mu\left(M_j\right) = \int_{-\infty}^{\infty} g_{M_j} f\left(\vec{Z}\right) d\vec{Z} \tag{3.15}$$

and the variance by:

$$\nu_{M_j} = \nu\left(M_j\right) = \int_{-\infty}^{\infty} \left(g_{M_j} - \mu\left(M_j\right)\right)^2 f\left(\vec{Z}\right) d\vec{Z} \tag{3.16}$$

- when there is an implicit linear relationship ($g_{M_j}(\vec{Z})$) between each of the system's output value (M_j) and the input values ($\vec{Z}$) and if each probability distribution ($f(Z_i)$) corresponding to each single input (Z_i) is normal and statistically independent from each other, then, also the outputs are distributed in a normal way and you may get accurate probabilistic information numerically, for each output of the system.

However, often in real cases of energy systems it is almost impossible to find an explicit relationship ($g_{M_j}\vec{Z}$) between each output of the system (M_j) and each input (Z_i), and sometimes it is even more complicated to solve the integrals in equations 3.15 and 3.16. Furthermore, to recall an implicit linear relationship affects the quality of the model used to represent the real system, causing a loss of information that may not be desirable. Therefore, two other possible cases could be considered using the RSA method coupled with the development in the Taylor series of the first order, to obtain approximate probabilistic information for each of the outputs of the system:

- if only the mean and variance are known for each input variable (Z_i), (that is not the exact PDF), and have an implicit relationship ($g_{M_j}(\vec{Z})$) between the outputs of the system (M_j) and inputs ($\vec{Z}$), then the approximate values of the mean and variance of each output can be estimated using the expansion in Taylor series centered around the average value μ_{Z_i} of the inputs, namely:

$$\begin{aligned} M_j \simeq g_{M_j}\left(\mu_{\vec{Z}}\right) + \sum_{i=1}^{n} \left(Z_i - \mu_{Z_i}\right) \left.\frac{\partial g_{M_j}\left(\mu_{\vec{Z}}\right)}{\partial Z_i}\right|_{\mu_{Z_i}} \\ + \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{n} \left(Z_i - \mu_{Z_i}\right)\left(Z_k - \mu_{Z_k}\right) \left.\frac{\partial^2 g_{M_j}\left(\mu_{\vec{Z}}\right)}{\partial Z_i \partial Z_k}\right|_{\mu_{Z_{ik}}} + \ldots \end{aligned} \tag{3.17}$$

  However, it is not possible to determine the exact PDF ($f(M_j)$) corresponding to each of the system output.

- If there is an implicit or explicit relationship between the outputs of the system and inputs , and also the PDF of each input of the system ($f(Z_i)$) is a normal distribution, statistically independent of each other, then the approximate values of the mean and of the variance of each output can be estimated using the Taylor series expansions centered around the average value of the inputs (see equation 3.18). Also in this case it is not possible to determine the exact PDF ($f(M_j)$) corresponding to each output of the system.

For both of these two cases, the mean and variance of each output approximated to the first order can be calculated as follows:

$$\mu_{M_j} = \mu\left(M_j\right) \simeq g_{M_j}\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \mu_{Z_n},\right) \tag{3.18}$$

$$\nu_{M_j} = \nu\left(M_j\right) \simeq \sum_{i=1}^{n}\left(\frac{\partial g_{M_j}}{\partial Z_i}\right)^2 \nu\left(Z_i\right) + \sum_{i=1}^{n}\sum_{k=1}^{n}\frac{\partial g_{M_j}}{\partial Z_i}\frac{\partial g_{M_j}}{\partial Z_k}\chi\left(Z_i, Z_k\right) \tag{3.19}$$

where the term $\chi\left(Z_i, Z_k\right)$ is the covariance of Z_i and Z_k. Assuming that $f(Z_i)$ are not related to each other, i.e. they are statistically independent, the variance can be expressed in a reduced form:

$$\nu_{M_j} = \nu\left(M_j\right) \simeq \sum_{i=1}^{n}\left(\frac{\partial g_{M_j}}{\partial Z_i}\right)^2 \nu(Z_i) \tag{3.20}$$

The square root of the variance represents the standard deviation of M_j, indicated with σ_{M_j}.
If a higher precision of the results is needed, the second order mean can be calculated with the following expression:

$$\mu_{M_j} = \mu\left(M_j\right) \simeq g_{M_j}\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \mu_{Z_n},\right) + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial^2 g_{M_j}}{\partial Z_i^2}\right)\nu(Z_i) \tag{3.21}$$

To find the variance of the second order must be known in the moments of the third and fourth order of all variables in input $\vec{Z}$, but this rarely happens. Thus, for practical purposes, the variance of the first order (equation 3.19) and the mean of the second order (equation 3.21).
The derivative $\partial g_{M_j}/\partial Z_i$ that appears in some of the above equations, is called **sensitivity of the response of the system** (system response sensitivity) for the output M_j associated to Z_i. An important advantage of RSA method is its capacity to estimate the impact of each single input uncertainty on the monitored outputs at the same time of the probabilistic analysis. This is done through the parameter called sensitivity, defined as the first derivative of outputs with respect to the input variable, that can be non-dimensionalized as it follows:

$$Sensitivity = \frac{\partial g_{M_j}}{\partial Z_i}\frac{Z_{i,nom}}{g_{M_{j,nom}}} \tag{3.22}$$

In this way, it is possible to know how much influence each input can have on the monitored parameters; the higher is this value the higher is the influence.
If there are no explicit relationships between the responses and inputs of the system, the partial derivatives with respect to the input variables of these functions cannot be determined analytically. However, these derivatives can be determined numerically using the finite difference schemes. For example, the j-th output can be written, using the mean of the i-th input variable, as below (equations 3.23,3.24).

$$M_{ji}^{+} = g\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \left(\mu_{Z_i} + \delta\right), \ldots\right) \tag{3.23}$$

$$M_{ji}^{-} = g\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \left(\mu_{Z_i} - \delta\right), \ldots\right) \tag{3.24}$$

Using the scheme of the second order centred finite difference, the sensitivity of the system response for M_j is expressed as:

$$\frac{\partial g_{M_j}}{\partial Z_i} = \frac{M_{ji}^{+} - M_{ji}^{-}}{2\delta} + o\left(\delta^2\right) \tag{3.25}$$

where the term $o\left(\delta^2\right)$ is called truncation error of the second order. If this error is not relevant to the simulation, the equation can be approximated considering only the first term and then the variance of the first order becomes:

$$\nu_{M_j} \simeq \sum_{i=1}^{n} \left(\frac{M_{ji}^{+} - M_{ji}^{-}}{2\delta}\right)^2 \nu(Z_i) \tag{3.26}$$

where $\nu(Z_i)$ is the variance of the input Z_i.
Note that the perturbation step must be sufficiently small to minimize the truncation error, but also high enough to avoid the sensitivity related to the error simulation. If the truncation error is large or the step is too small, it should be used the scheme of the centred finite differences of the fourth order, which presents a fourth order truncation error. In this case, the sensitivity of the response becomes:

$$\frac{\partial g_{M_j}}{\partial Z_i} = \frac{M_{ji}^{--} - 8M_{ji}^{-} + 8M_{ji}^{+} - M_{ji}^{++}}{12\delta} \tag{3.27}$$

While the second derivative always expressed with fourth order centred scheme becomes:

$$\frac{\partial^2 g_{M_j}}{\partial Z_i^2} = \frac{16M_{ji}^{-} - M_{ji}^{--} - 30M_{ji} + 16M_{ji}^{+} - M_{ji}^{++}}{12\delta} \tag{3.28}$$

where

$$M_{ji}^{++} = g\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \left(\mu_{Z_i} + 2\delta\right), \ldots\right) \tag{3.29}$$

$$M_{ji}^{--} = g\left(\mu_{Z_1}, \mu_{Z_2}, \ldots, \left(\mu_{Z_i} - 2\delta\right), \ldots\right) \tag{3.30}$$

To summarize the main features of the RSA method, it can be stated that:

- the RSA requires information only on statistical parameters such as mean and

variance of the input variables, and do not need their full distributions (normal, gamma, Poisson, etc.)

- the properties of the statistical results are obtained directly from Taylor series expansions
- care must be taken if the input variables are statistically independent or not, as this is detected in terms of precision of the result and on the choice of the type of approximate relationship to use. In fact, if the input are not independent, the covariance between them must be considered in the analysis
- the use of too low order schemes or too coarse approximations can cause errors especially for highly nonlinear problems
- the RSA method is directly applicable to the analysis of large-scale systems for which the response has a linear behaviour even if the function of the response is not explicitly known.

In conclusion, the biggest advantage of the RSA method is that uncertainty of the system output variables is evaluated with a small number of calculations, while the simulation with the Monte Carlo method requires an extremely high number of sampling simulations. The number of the RSA calculations is proportional to the number of variables. This is therefore the most practical methodology for the analysis and large-scale optimization of a complex energy system.

3.2.3 Polynomial Chaos

Among the approximated methods for uncertainty quantification, the Polynomial Chaos (PC) is the most popular. It based on the work by Wiener [28], which was originally concerned with the stochastic processes of Gaussian random variables. The topic of polynomial chaos has received some attention in the last twenty years as a mean to efficiently estimate model outcomes based on known stochastic processes. A PC represents a stochastic process as a polynomial expansion, with corresponding expansion coefficients, over random variables of known distributions. This approach can exhibit very fast convergence if the observables depend smoothly on the random parameters. In addition, moment estimation and sensitivity analysis (Sobol indices) can be extracted without significant additional costs from the PC expansion. In the PC framework, both intrusive and non-intrusive PC methods are available [78].
Intrusive methods inject the PC into the model governing equations and solve for all coefficients simultaneously via a Galerkin Projection. Hence, the intrusive approach requires the governing equations to be rewritten since all dependent variables and random parameters in the governing equations are replaced with their polynomial chaos expansions. Non-linear governing equations are especially challenging, often requiring special treatment of the Galerkin projection [79]. Although straightforward in theory, an intrusive formulation for complex problems can be relatively difficult, expensive, and time consuming to implement. This means altering the source code, which is used for the computations. In some cases this can be a disadvantage (e.g. for well validated industrial CFD codes, where any extension has a risk of introducing errors). However, it can result in a sensible performance increment if used

to compute with multiple uncertainties since it provides immediate access to the coefficients through a single (though much larger) model solution but the structure of the stochastic problem may require different solvers from those of the deterministic model [80]. Because they require a complete reformulation of the model governing parameters, intrusive methods are usually limited to simpler models. While useful to gain insight into the response of a stochastic model, such reformulation makes intrusive projection prohibitive for large, legacy solvers.
To overcome such inconveniences associated with the intrusive approach, Non-Intrusive Polynomial Chaos (NIPC) formulations have been considered for uncertainty propagation. The main objective of the NIPC method is to obtain the polynomial coefficients without making any modifications to the deterministic code. This approach treats the deterministic code as a "black-box" and approximates the polynomial coefficients with formulas based on deterministic code evaluations. The way to gain information about the system is by running simulation with some specific sampling data points. The "ideal" non-intrusive method would predict the polynomial coefficients with minimum number of deterministic evaluations at the desired accuracy level for a given stochastic problem. For this reason, the non-intrusive approach is much more common for engineering applications as it is simpler to apply and it is the approach used in this analysis. The reason for this is that it is much easier to run computations with a readily available commercial code than it is to rewrite the governing equations in the source code to include intrusive PC.

Generalized polynomial chaos

Polynomial Chaos, also called "Wiener Chaos expansion", is a spectral method to propagate uncertainty in a system, when there is aleatory uncertainty in the system parameters. An important aspect of spectral representation of uncertainty is that one may decompose a random function (or variable) into separable deterministic and stochastic components. The stochastic response output can be approximated by a series of orthogonal polynomials basis from Askey scheme [28] associate with random inputs. Xiu and Karniadakis [34, 81] introduce the Generalized Polynomial Chaos (gPC), also called Askey Chaos, proposing a more general framework of polynomial chaos employing the classes of orthogonal polynomials of the Askey scheme, including the Hermite chaos as a subset. Due to the closed connection between the orthogonal polynomials in the Askey scheme to the probability density function of certain random distributions, the generalized polynomial chaos allows to represent many non-Gaussian stochastic processes, including some discrete processes. It has the potential to achieve a significant reduction in computational cost (number of evaluations) with respect to traditional techniques such as Monte Carlo approaches. Regardless the distinction between intrusive and non-intrusive method, all Polynomial Chaos methods works based on the same principle. An approximation of the model is constructed using an orthogonal set of polynomials, which serve as basis functions for an N-dimensional parameter space. For example, any random variable (i.e., R) can be written as:

$$R\left(\vec{x},\vec{\xi}\right) = \sum_{j=0}^{\infty} a_j(\vec{x})\psi_j(\vec{\xi}) \tag{3.31}$$

where $a_j(\vec{x})$ and $\psi_j(\vec{\xi})$ are the polynomial expansion coefficients and random basis function corresponding to the j-th node, respectively. In the most general case, $a_j(\vec{x})$ can be a function of deterministic independent variable vector $\vec{x}$ and the n-dimensional standard random variable vector $\vec{\xi} = (\xi_1, \ldots, \xi_n)$. Each of the $\psi_j(\vec{\xi})$ are multivariate polynomials which involve products of the one-dimensional polynomials.
In theory, the polynomial chaos expansion given by Equation 3.31 should include infinite number of terms; however, in practice the infinite expansion can be truncated at a finite number of random variables (n) and a finite expansion order (p):

$$R\left(\vec{x}, \vec{\xi}\right) \simeq \sum_{j=0}^{p} a_j(\vec{x})\psi_j(\vec{\xi}) \tag{3.32}$$

where the total number of terms N in a complete polynomial chaos expansion of arbitrary order p for a response function involving n uncertain input variables is given by:

$$N = 1 + P = 1 + \sum_{s=1}^{p} \frac{1}{s!} \prod_{r=0}^{s-1} (n + r) = \frac{(n + p)!}{n!p!} \tag{3.33}$$

The polynomial chaos expansion obtained with this truncation approach can be labelled as "total-order expansion".
The corresponding p-th degree polynomial chaos approximation for a function $R(x)$ of d random variables $\vec{\xi} = (\xi_1, \ldots, \xi_n)$ with PDF(x) is:

$$R_d^q(x) = \sum_{m=1}^{Q} a_m \psi_m(\vec{\xi}) \tag{3.34}$$

$$Q = \binom{q + d}{q} \tag{3.35}$$

with expansion coefficients and multivariate orthonormal polynomials $\psi_m(\vec{\xi})$. Here, the summations are over all possible combinations of the multi-index $m = (m_1, \ldots, m_d)$ such that $|m| = \sum m \leq q$. The $\psi_m(\vec{\xi})$ are selected such that they are orthonormal with respect to the weight function (x) in domain Ω, namely:

$$\int_{\Omega} \psi_i(x)\psi_j(x)\pi(x)\, dx = \delta_{ij} \tag{3.36}$$

where δ_{ij} is the Kronecker delta. This orthonormality allows the expansion coefficient to be determined by:

$$a_m = \int_{\Omega} f_x P\psi_m(x)\pi(x)\, dx \tag{3.37}$$

For gPC, polynomials are determined which best match the distribution of the non-Gaussian input random variables [81].
The basis function ideally takes the form of multi-dimensional Hermite Polynomial to span the n-dimensional random space when the input uncertainty is Gaussian (unbounded), which was first used by Wiener [28] in his original work of polynomial chaos. To extend the application of the polynomial chaos theory to the propagation of continuous nonnormal input uncertainty distributions, Xiu and Karniadakis [34] used a set of polynomials known as the Askey scheme to obtain the "Wiener-Askey Generalized Polynomial Chaos". Table 3.3 shows the correspondence between the type of Wiener-Askey polynomial chaos and the type of random inputs.

TABLE 3.3: Relationship between standard forms of continuous probability distributions and Askey scheme of continuous orthogonal polynomials

	Random inputs	**Wiener-Askey chaos**	**Support**
Continuous	Gaussian	Hermite-chaos	$(-\infty, \infty)$
	Gamma	Laguerre-chaos	$[0, \infty]$
	Beta	Jacobi-chaos	$[a, b]$
	Uniform	Legendre-chaos	$[a, b]$
Discrete	Poisson	Charlier-chaos	$(0, 1, \ldots, N)$
	Binomial	Krawtchouk-chaos	$(0, 1, \ldots, N)$
	Negative binomial	Meixner-chaos	$(0, 1, \ldots, N)$
	Hypergeometric	Hahn-chaos	$(0, 1, \ldots, N)$

The generalized polynomial chaos approach can be applied to the propagation of any independent random variable included in the Askey scheme. The detailed information on polynomial chaos expansions can be found in Cuneo, Traverso and Shahpar [23] Najm [82] and Eldred and Burkardt [83].
The convergence of the polynomial chaos expansion is ensured by the theorem of Cameron and Martin and the generalized version respectively. In several papers, the convergence rate depending on the order of the polynomials used in the polynomial chaos expansion is numerically examined [84, 81]. The total number of expansion terms increases fast for large dimensional problems, thus even with exponential convergence, the polynomial chaos method is computationally effective, only if the number of input random variables is not too large. Techniques reducing the number of random variables can circumvent this limitation. Methods investigating the influence of the individual random variables on the output have the potential of reducing the dimension of the probability space [83, 85]. Determining the expansion coefficients in equation 3.33 involves the potentially timeconsuming task of evaluating sufficient forward problems, $f(x)$, to accurately estimate these coefficients. However, once these coefficients have been computed, any future function estimations only require the evaluation of 3.33. Thus, the potentially large number of forward-problem evaluations needed to quantify the distribution of $f(x)$ via Monte Carlo sampling is replaced by the potentially much smaller number of evaluations needed to precompute the coefficients for the PC expansion, from which a Monte Carlo analysis can be inexpensively performed. In order for this to pay off, we need a low-degree expansion and accurate computation of the coefficients in 3.35 with minimal forward evaluations.

3.3 Analysis of Variance

Analysis of Variance (ANOVA) is a statistical technique that is used to check if three or more groups are statistically different from each other. ANOVA checks the impact of one or more factors by comparing the means of different samples [86].
In order to do that, ANOVA uses a so-called Null hypothesis and an Alternate hypothesis. The **null hypothesis (H_0)** in ANOVA is valid when all the sample means are equal, or they don't have any significant difference. Thus, they can be considered as a part of a larger set of the population. On the other hand, the **alternate hypothesis (H_1)** is valid when at least one of the sample means is different from the rest of the sample means. In mathematical form, they can be represented as:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k \tag{3.38}$$

$$H_1 : \mu_l \neq \mu_m \tag{3.39}$$

where μ is the group mean and k indicates the number of groups considered. In other words, the null hypothesis states that all the sample means are equal or the factor did not have any significant effect on the results. Whereas, the alternate hypothesis states that at least one of the sample means is different from another. But why should we use an ANOVA test when we could conduct a series of t-test? [87, 88]
Let us consider that we want to compare the level of satisfaction of customers for four difference restaurants, in order to understand if there is a difference of satisfaction across them. The satisfaction scores for a sample of customers for each restaurant are recorded as in table 3.4

TABLE 3.4: Sastisfaction scores of customers

Restaurant 1	Restaurant 2	Restaurant 3	Restaurant 4
3.2	4.2	5.4	4.5
3.5	3.7	4.6	3.8
2.7	3.4	4.0	4.1
4.1	4.3	5.3	3.1
3.1	3.9	4.7	4.2
3.7	4.1	4.2	3.4
4.2	3.1	4.9	4.2
3.6	4.5	4.7	4.5

One of the mean advantages of ANOVA is that it allows to test the means of three or more populations at once, while t-test would require to run six indipendent samples t-test comparing two restaurants one another for each sample (i.e. restaurant 1 vs. restaurant 2, restaurant vs. restaurant 3, etc.). This process would be long, but with the help of a computer would not be too hard to perform. However, doing so would lead to other drawbacks, as when more than one t-test is run, each at its own level of significance, the probability of making one or more Type I errors multiplies

exponentially. A Type I error occurs when we reject the null hypothesis when we should not. The level of significance, α, is the probability of a Type I error in a single test. So, for a single t-test in our example, with an α of 0.05, we would have a Type I error probability of 5%. When testing more than one pair of samples, the probability of making at least one Type I error is:

$$1 - (1 - \alpha)^c \tag{3.40}$$

where α is the level of significance for each t-test and c is the number of independent t-tests. Considering the example reported in table 3.4, the probability of committing a Type I error would be of 26.5% instead of 5%. For this reason the ANOVA test is highly suggested when comparing three or more groups.
Before we can use the one-way ANOVA, where one-way means there it has just one main grouping factor, we must see if we satisfy some assumptions,such as:

1. all observations are independent of one another and randomly selected from the population which they represent.
2. The population at each factor level is approximately normal.
3. The variances for each factor level are approximately equal to one another

With the ANOVA method, we are actually analyzing the total variation of the scores, including the variation of the scores within the groups and the variation between the group means. Since we are interested in two different types of variation, we first calculate each type of variation independently and then calculate the ratio between the two, called F-value. ANOVA has its own distribution that we need to use, called an F-distribution to set our critical values and test our hypothesis. The F-distribution relies on degrees of freedom. Since the F-value is actually a ratio of two different sources of variance, we'll need two different degrees of freedom. When using the ANOVA method, we are testing the null hypothesis that the means and the variances of our samples are equal. When we conduct a hypothesis test, we are testing the probability of obtaining an extreme F-statistic by chance. If we reject the null hypothesis that the means and variances of the samples are equal, and then we are saying that the difference that we see could not have happened just by chance. To test a hypothesis using the ANOVA method, there are several steps that we need to take. To help us in completing those steps, we need to employ the so-called ANOVA table (table 3.5) [89].

TABLE 3.5: ANOVA table

Source	**SS**	d_f	**MS**	**F**	$F_{critical}$
Between					
Within					
Total					

The left column of table 3.5 lists where the variation in the test is coming from: between the groups, within the groups, or all the variance for all the observations

(Total). Finally, the steps which need to be taken to perform the one-way ANOVA are the following [86, 90]:

1. **Calculate the total sum of squares** SS_T. Computed as the difference between each score and the grand mean, which is the mean of all the scores disregarding the group:

$$SS_T = \sum \left(y - \overline{y}^2\right) \tag{3.41}$$

where y is the value of each observation and $\overline{y}$ is the grand mean. For the example reported in table 3.4, $SS_T = 12.965$.

2. **Calculate the sum of squares between** SS_B as:

$$SS_B = \sum n_k \left(\overline{y}_k - \overline{y}\right)^2 \tag{3.42}$$

where k is the number of groups, n_k the number of scores in group k and $\overline{y}_k$ the mean of group k. Considering table 3.4, where $k = 4$ and $n_k = 8$, SS_B would be then equal to 6.166.

3. **Calculate the sum of squares within groups** SS_W as:

$$SS_W = SS_T - SS_B \tag{3.43}$$

4. **Compute the degress of freedom for the test**:

$$d_{f,total} = N - 1 \tag{3.44}$$

$$d_{f,between} = k - 1 \tag{3.45}$$

$$d_{f,within} = N - k \tag{3.46}$$

5. **Calculate the Mean Squares Between (**MS_B**) and Mean Square Within (**MS_W**)** using the relationships below:

$$MS_B = \frac{SS_B}{d_{f,between}} \tag{3.47}$$

$$MS_W = \frac{SSW}{d_{f,within}} \tag{3.48}$$

6. **Calculate the F statistic** as:

$$F = \frac{MS_B}{MS_W} \tag{3.49}$$

7. **Fill in the ANOVA table**, which for our example would become (table 3.6)

8. **Find** $F_{critical}$ by considering the degrees of freedom between and within (i.e. 3,28) and the α value (i.e. 0.05). This can be done by consulting proper tables, such as [91]. In our case we would have an $F_{critical}$ of 2.947.

TABLE 3.6: ANOVA table

Source	SS	d_f	MS	F	$F_{critical}$
Between	6.166	3	2.055	8.457	
Within	6.799	28	0.243		
Total	12.965	31			

9. **Interpret the results of the hypothesis test**. In ANOVA, the last step is to decide whether to reject the null hypothesis and then provide clarification about what that decision means. In our example of the restaurants, the F-value from the ANOVA test is greater than the F-critical value, so we would reject the Null Hypothesis. We could then conclude that the average customer satisfaction scores of the four restaurants are not equal to one another – at least one of them is different from the others.
   Another measure for ANOVA is the p-value. If the p-value is less than the alpha level selected, we reject the Null Hypothesis [92].

3.4 Design of Experiment

Response Surface Methodology (RSM) is a tool that was introduced in the early 1950s by Box and Wilson (1951).It consists of a group of mathematical and statistical techniques that can be used to define the relationships between independent input variables and interested outputs, which are called response variables.
An important aspect of RSM is the Design of Experiment [76], usually abbreviated as DoE. These strategies were originally developed for the model fitting of physical experiments, but can also be applied to numerical experiments. The objective of DoE is the selection of the points where the response should be evaluated for guiding the choice of the experiments to be performed in an efficient way.
Usually, data subject to experimental error (noise) are involved, and the results can be significantly affected by noise. Thus, it is better to analyze the data with appropriate statistical methods. The basic principles of statistical methods in experimental design are replication, randomization, and blocking. Replication is the repetition of the experiment in order to obtain a more precise result (sample mean value) and to estimate the experimental error (sample standard deviation). Randomization refers to the random order in which the runs of the experiment are to be performed. In this way, the conditions in one run neither depend on the conditions of the previous run nor predict the conditions in the subsequent runs. Blocking aims at isolating a known systematic bias effect and prevent it from obscuring the main effects. This is achieved by arranging the experiments in groups that are similar to one another. In this way, the sources of variability are reduced and the precision is improved [93].
A particular combination of runs defines an experimental design. The possible settings of each independent variable in the N-dimensional space are called levels. The number of levels usually is the same for all the variables, however some DoE techniques allow to use a different number of levels for each variable. In experimental design, the objective function and the set of the experiments to be performed are called response variable and sample space respectively. Several techniques to perform a proper DoE have been developed along the years, however in the following

only some of them will be described, with a particular focus on the ones which have been used by the author in his works.

3.4.1 Factorial Design

The term "*factorial design*" is thought to have been introduced for the first time in statistical literature by Ronald Fisher [94] in 1935. As a general definition, a factorial design consists of two or more factors, each with dicrete possible levels. If all the possibile combinations of these levels across all such factors are considered, then the design is defined as "*full factorial*". On the other hand, if a chosen subset of the experimental runs of a full factorial design is used, then the design is called "*fractional factorial*". The subset is chosen so as to exploit the sparsity-of-effects principle, i.e. to expose information about the most important features of the problem studied, while using a fraction of the effort of a full factorial design in terms of experimental runs and resources.
The most simple form of the factorial design is the one which employs two levels per factor. The two levels are called high and low, and in literature are often in indicated with "h" and "l", or "$+1$" and "-1" respectively. If there are k factors, each at 2 levels, a full factorial design will be of 2^k runs. To give a better understanding of this method, let us consider a full factorial design with three factors and two levels per factor. It becomes immediately clear that the number of runs needed to perform a full factorial design will be of $2^3 = 8$. Graphically, it is possible to denote this design with a cube as shown in figure 3.5.

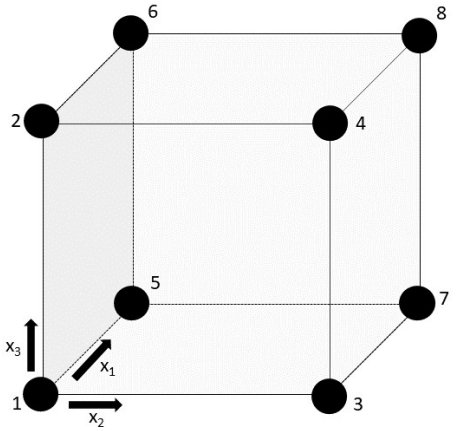


FIGURE 3.5: Graphical representation of a 2^3 full factorial design

The numbers in the corners indicates the standard order of runs, as enlisted in the following table:

We define also the main interaction M of a variable x as the difference between the average response variable at the high levels and the average response at the low level samples [93]. For the example reported in table 3.7, for x_1 we would have:

$$M_{x_1} = \frac{y_{h,l,l} + y_{h,l,h} + y_{h,h,l} + y_{h,h,h}}{4} - \frac{y_{l,l,l} + y_{l,l,h} + y_{l,h,l} + y_{l,h,h}}{4} \tag{3.50}$$

TABLE 3.7: Standard order of a 2^3 full factorial design

Experiment number	**Factor level** x_1	x_2	x_3	**Response variable**	**Two and three factors interactions** $x_1 \cdot x_2$	$x_1 \cdot x_3$	$x_2 \cdot x_3$	$x_1 \cdot x_2 \cdot x_3$
1	-1	-1	-1	$y_{l,l,l}$	+1	+1	+1	-1
2	-1	-1	+1	$y_{l,l,h}$	+1	-1	-1	+1
3	-1	+1	-1	$y_{l,h,l}$	-1	+1	-1	+1
4	-1	+1	+1	$y_{l,h,h}$	-1	-1	+1	-1
5	+1	-1	-1	$y_{h,l,l}$	-1	-1	+1	+1
6	+1	-1	+1	$y_{h,l,h}$	-1	+1	-1	-1
7	+1	+1	-1	$y_{h,h,l}$	+1	-1	-1	-1
8	+1	+1	+1	$y_{h,h,h}$	+1	+1	+1	+1

and similar expressions could be derived for x_2 and x_3. The *interaction effect* of two or more factors is defined similarly as the difference between the average responses at the high level and at the low level in the interaction column. The two-factors interaction effect between x_1 and x_2 as expressed in table 3.7 would be:

$$M_{x_1,x_2} = \frac{y_{l,l,l} + y_{l,l,h} + y_{h,h,l} + y_{h,h,h}}{4} - \frac{y_{l,h,l} + y_{l,h,h} + y_{h,l,l} + y_{h,l,h}}{4} \tag{3.51}$$

The main interaction and the interaction effect give a quantitative estimation of the influence of the factors on the response variable. The number of main and interaction effects in a 2^k full factorial design is $2^k - 1$, which correspond also to the degree of freedom of such design.
The full factorial design makes a very efficient use of the data and is particularly valuable because it does not overshadow the effect of some of the parameters, instead it is possible to evaluate the interaction effects clearly. On the other hand, the sample size grows exponentially with the number of parameters and the number of levels, so it can lead to an unfeasible number of experimental runs. Therefore, a fractional factorial design can be more useful in some cases, considering however that the choice of the proper subset of runs to perform can be a critical issue.

3.4.2 Central Composite Design

A central composite design is a 2^k full factorial to which are added the central points and the star points. The star points are those points in which all the parameters but one are set at the mean, or central, level. The points at the center of the experimental domain and the star points outside this domain make it possible to estimate the curvature of the response surface. The distance of the star point from the central point is called α, and depending on its value three main different types of central composite design can be identified [95]:

- **Central Composite Circumscribed (CCC)** (figure 3.6a) when all the samples are placed on a hypersphere centered in the central point and the star points

are outside the design space of a regular full factorial design. Five levels for each factor are required. This method allows the highest quality predictions and is easy to create from a full factorial experiment, however the star points can be unfeasible points or give unreasonable results;

- **Central Composite Inscribed (CCI)** (figure 3.6b) if a sampling like the central composite circumscribed is desired, but the limits specified for the levels cannot be violated. This design can guarantee also high quality predictions, even if lower than CCC, and eliminates the main disadvantage of the CCC linked to the positioning of the star points outside the design space. However, this design cannot be directly derived from a full factorial design;
- **Central Composite Face Centered (CCF)** (figure 3.6c) where the star points are located on the faces of the full factorial design. This method requires three level for each factor, gives mostly high quality predictions and is easy to create from a full factorial design. However it can give a poor quality prediction of all the pure quadratic effects.

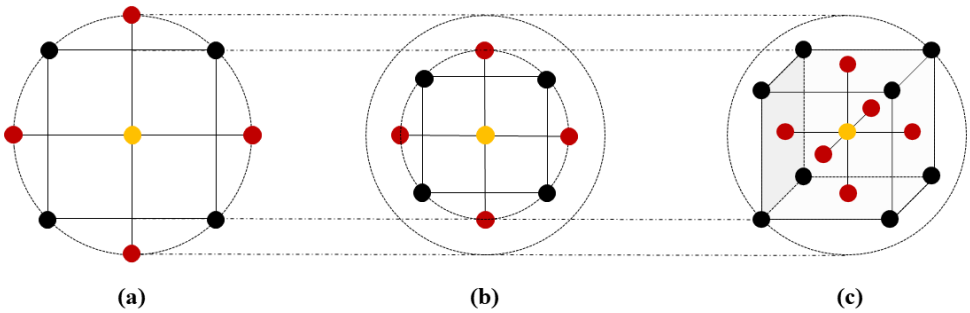


FIGURE 3.6: Examples of: (a) 2^2 CCC, (b) 2^2 CCI, (c) 2^3 CCF

When considering k parameters, 2k star points and one central point are added to the 2^k full factorial, bringing the sample size for the central composite design to $2^k + 2k + 1$.

3.4.3 Box-Behnken Design

Box-Behnken is an incomplete three level design built combining two-level factorial design with incomplete block design and is usually used to fit second-order model regression [96]. Considering a three factor design, the graphical representation of Box-Behnken can be seen in two forms: a cube with a central point (indicated in orange) and the middle points of the edges (figure 3.7a), or as three interlocking 2^2 factorial design with a central point (figure 3.7b).

For this method, when considering k parameters, then sample size is equal to $2k(k-1) + n_0$ where n_0 correspond to the number of central points which in the case reported in figure 3.7 is equal to 1, as it is the same for all the three interlocking 2^2 factorial designs. Box-Behnken experimets are particularly useful if some boundary areas of the design region are unfeasible, as it does not take into account the corners of the design space. On the other hand it can have a poor prediction in the corner of the design space, which could be a critical issue for some cases.

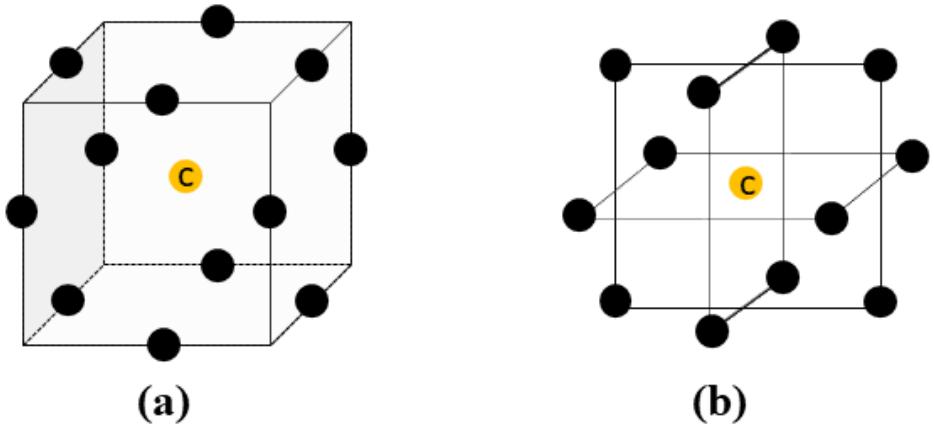


FIGURE 3.7: Box-Behnken graphical representation: (a) cube, (b) three interlocking 2^2 factorial designs

3.4.4 Taguchi

The Taguchi method was developed by Dr. Genichi Taguchi in Japan [97] to improve the off-line quality control, ensuring good performance in the design stage of products or processes. For Taguchi *"quality is measured as the total loss to society caused by a product"*, so if the final product does not meet the standard set it could cause a damage to the society. The idea of Taguchi was to find the best values of the controllable factors to make the problem less sensitive to the variations in the uncontrollable factors. In particular, the Taguchi process can be divided in the following steps:

1. **Problem identification**:
   - locate the source of the problem, not just the symptom.
2. **Brainstorming**:
   - identify the critical variables (factors) for the quality of the product or process analyzed:
     - **control factors**: variables under management control
     - **noise factors**: uncontrollable factors
   - define the factor levels and identify the possible interactions between them
   - determine the objective of the experiment:
     - **smaller-the-better**: to keep the level of defectives as close to zero as possible (minimize the response variable)
     - **larger-the-better**: maximize the number of units per time unit or lot without defects (maximize the response variable)
     - **nominal-the-best**: outcome as close to target as possible.
3. **Experimental design**:
   - determine the design for control and noise factor considering the factor levels and objectives decided during the brainstorming.
4. **Analysis**:
   - figure out the best levels in terms of robustness and evaluate the outputs

When considering the two set of factors (control and noise), two different orthogonal designs are chosen. The design chosen for the controllable variable is called

inner array, while the one chosen for the noise variables is called *outer array*. The combination of the inner and the outer arrays gives the crossed array which is the list of all the samples scheduled by the Taguchi method. By combination we mean that the full set of experiment for the outer array is performed for each sample in the inner array.
To give the reader an example of the Taguchi method, we will consider a problem with five parameters, three of which are controllable and two uncontrollable, and a two-level factorial design both for inner and for outer arrays. Then, we must perform a full 2^2 factorial design (outer array) for each sample of the 2^3 inner array, which graphically is represented as in figure 3.8.

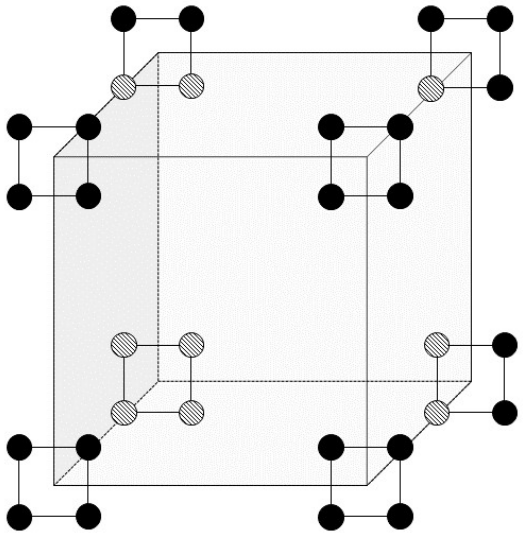


FIGURE 3.8: Taguchi graphical representation for 3 controllable factors and 2 noise factors

According to the Taguchi method, the inner and the outer arrays are to be chosen from a list of published orthogonal arrays [98], which are dependent on the number of variables and levels considered.

3.4.5 Latin hypercube

Latin hypercube is a space-filling design technique, in which the design space is subdivided into an orthogal grid with N elements of the same length per parameter. Within the grid N sub-volumes are located so that along each row and column of the grid only one sub-volume is chosen. In figure 3.9 an example of latin hypercube is shown, where the black painted square represent the chosen sub-volumes. Inside each sub-volume a sample is randomly chosen [99, 100].
It is important to choose the sub-volumes in order to have no spurious correlations between the dimensions or,which is almost equivalent, in order to spread the samples all over the design space. For instance, a set of samples along the design space diagonal would satisfy the requirements of a latin hypercube DOE, although it would showa strong correlation between the dimensions andwould leave most of the design space unexplored [93, 101].

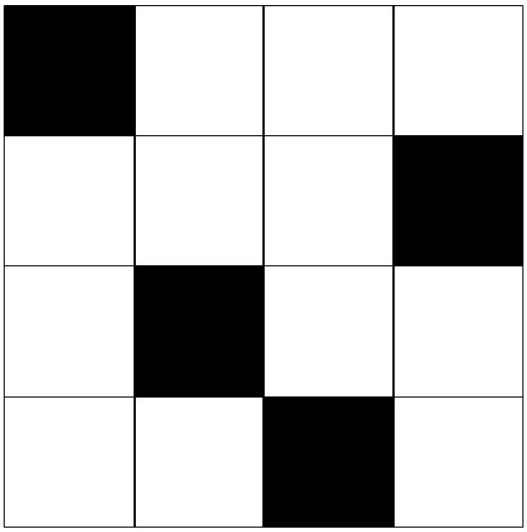


FIGURE 3.9: Example of latin hypercube design with $k = 2$ and $N = 4$

Let us assume a generic case where we have k parameters and N samples. In order to compute a set of Latin hypercube samples two matrices $Q_{N\times k}$ and $R_{N\times k}$ are built [102]. The colums of the $Q_{N\times k}$ are random permutations of the integers between 1 and N, while $R_{N\times k}$ matrix contains random values uniformly distributed in $[0, 1]$. Assuming each parameter in the range $[0, 1]$, we can create the sampling map matrix $S_{N\times k}$ containing the samples in its row as:

$$S = \frac{1}{N}(Q - R) \tag{3.52}$$

To reduce auto-correlations we operate on Q, creating the matrix $Y_{N\times k}$ through the normal Gaussian cumulative distribution function D_{norm}, so that the elements of Y are in the range $[-1, 1]$:

$$y_{i,j} = D_{norm}^{-1}\left(\frac{q_{i,j}}{N+1}\right) \tag{3.53}$$

Then the covariance matrix of Y is computed and Choleski decomposed

$$C = covY = LL^T \tag{3.54}$$

so that its elements are:

$$c_{i,j} = \frac{1}{N}\sum_{l=1}^{N}(y_{l,i} - \mu_i)(y_{l,j} - \mu_j) \tag{3.55}$$

where μ_i is the average of the values in the i-th column of Y. The Choleski decomposition requires C to be positive definite [103, 104]. For the way the matrix is

built this is guaranteed if $N > k$. A new matrix Y^* is computed so that

$$Y^* = Y\left(L^{-1}\right)^T \tag{3.56}$$

and the ranks of the elements of the columns of Y^* become the elements in the columns of the matrix Q^* which is used in place of Q in order to compute the samples. In figure 3.10 shows an example of correlation reduction applied to a design with $k = 2$ and $N = 10$ [93].

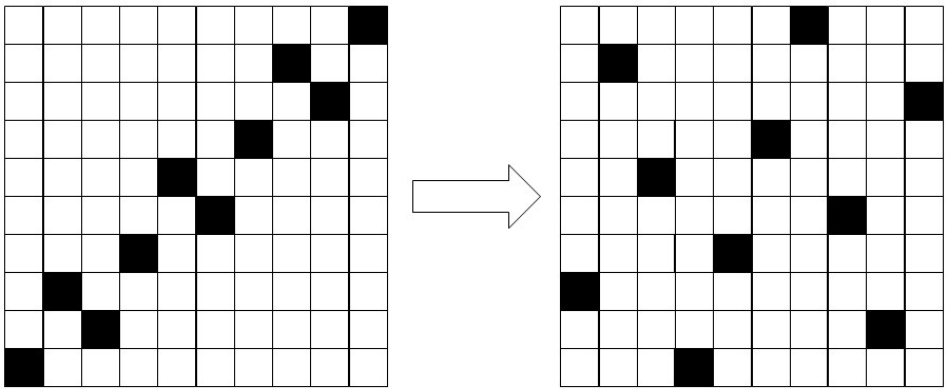


FIGURE 3.10: Example of correlation reduction in latin hypercube design with $k = 2$ and $N = 10$

3.4.6 Concluding remarks

Many DoE techniques are available, which have not been fully covered in this thesis as it was not the purpose of this work. The best choice depends on the nature of the problem to be investigated, as well as with the familiarity of the user with a certain technique which could simplify or harden the whole process. However, when deciding the DoE technique to be used it is important to consider:

- **the number of experiments** that can be afforded, considering the time required for a single experiment.
- **the number of parameters k** to consider. Unless are able to perform very high number of experiments, it is good practice to start from a preliminary study of the main effects in order to reduce k. Of course the choice of the parameters to be discarded can be a particularly delicate issue.
- **the number of levels L** for each parameters which mostly depends on the expected regularity of the response variable. the number of levels must be chosen carefully: it must be limited when possible, and it has to be kept higher if an irregular behaviour of the response variable is expected.
- **the aim of the DOE**: depending on the kind of insights that we want to have we should choose a different DoE technique. For example, if a more precise computation of the main and some interaction effects must be accounted for, a

fractional or a full factorial method is better. If the aim is to focus on a primary factor a latin square or a randomized complete block design would be suitable. If noise variables could influence significantly the problem a Taguchi method is suggested, even if a relatively cheap method also brings drawbacks. For RSM purposes, a Box-Behnken, a full factorial, a central composite, or a space filling technique has to be chosen [93].

3.5 Response Surface Methodology

Response Surface Methodology (RSM) is a tool that was introduced in the early 1950s by Box and Wilson (1951).It consists of a group of mathematical and statistical techniques that can be used to define the relationships between independent input variables and interested outputs, which are called response variables.
Such methodology helps to evaluate the effects of several parameter and establish the optimum conditions for the desired responses, through the creation of a mathematical model called metamodel [75]. Metamodels are developed to obtain a better understanding of the nature of the relationship between the input variables and the output variables of the system under investigation. The Response Surface (RS) represents the graphical perspective of the mathematical model [105].
For example, we can simplify the process of cooking a pizza by saying that the final result will be influenced by the cooking time x_1 and the temperature x_2. The dish will change its final taste under any combination of treatment x_1 and x_2. Therefore, time and temperature can vary continuously. When treatments are from a continuous range of values, then a Response Surface Methodology is useful for developing, improving, and optimizing the response variable. In this case, the final result of the pizza y is the response variable, and it can be defined as a function of time and temperature of cooking:

$$y = f(x_1, x_2) + \epsilon \tag{3.57}$$

The variables x_1 and x_2 are independent variables where the response y depends on them. The dependent variable y is a function of x_1, x_2, and the experimental error term, denoted as ϵ. The error term ϵ represents any measurement error on the response, as well as other type of variations not counted in f. It is a statistical error that is assumed to distribute normally with zero mean and variance s^2. In most RSM problems, the true response function f is unknown. In order to develop a proper approximation for f, the experimenter usually starts with a low-order polynomial in some small region. If the response can be defined by a linear function of independent variables, then the approximating function is a first-order model [106]. A first-order model with N independent variables can be expressed as:

$$y = \beta_0 + \sum_{i=1}^{N} (\beta_i \cdot x_i) + \epsilon \tag{3.58}$$

If instead the interaction between the the N indepedent variables needs to be taken into account, then the interaction model should be considered:

$$y = y = \beta_0 + \sum_{i=1}^{N}(\beta_i \cdot x_i) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\beta_{ij} \cdot x_i \cdot x_j) + \epsilon \tag{3.59}$$

and finally if there is a curvature in the response surface, a higher degree polynomial should be used. The approximating function with N variables is called a second-order model:

$$y = \beta_0 + \sum_{i=1}^{N}(\beta_i \cdot x_i) + \sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\beta_{ij} \cdot x_i \cdot x_j) + \sum_{i=1}^{N}(\beta_i i \cdot x_i^2) + \epsilon \tag{3.60}$$

where y is a response variable, x_i the factors or variables which have to be correlated, b_i the linear coefficients, β_0 the constant coefficient, β_{ii} the quadratic coefficients for variable i and β_{ij} the linear model coefficients for the interaction between variables i and j.
In general all RSM problems use either one or the mixture of the both of these models. In each model, the levels of each factor are independent of the levels of other factors. In order to get the most efficient result in the approximation of polynomials the proper experimental design must be used to collect data.
Therefore, RSM can be used for:

- locate, from the response surface, the area in which the optimum is expected to be: this allows to redefine the constraints over the input variables in order to shrink the design space in the neighbourhood of the optimum. The shrinked design space is then employed for the subsequent optimization process,
- create a metamodel to be used with an optimization algorithm, fully or partially replacing the experiments or the simulations. If used for a partial replacement, the metamodel can also be built directly, using the optimization data. From one side this means to build the RSM using potentially clustered data, as noted above; on the other side the fact that the response surface can be updated each time that data from a new simulation are made available may be very advantageous.

In general, response surface modelling includes the following steps:

1. data generation
2. model structure selection
3. parameter estimation
4. model validation

In the literature, a number of mathematical functions have been used to develop these metamodels [107, 108, 109, 110]. The data for building the response surface are generally collected employing a suitable experimental design. In the following subsections some of the main methodologies are exposed.

3.5.1 Least Squares Method

Least Square Method (LSM) is a method firstly developed by Gauss around 1795 and published several years later [111]. It is an approximated method in which the best fit of a data set is sought by tuning the β coefficients of a model function defined as $\hat{f}(x, \beta)$. We consider $\beta = [\beta_1, \beta_2, \ldots, \beta_m]^T$ the vector of m unknown coefficients to be found and $x = [x_1, x_2, 1dots, x_K]^T$ the vector of k input parameters. The data set consists of (x_i, y_i) pairs, with $i = 1, \ldots, N$, where x_i is the input parameter of the i-th simulation, whose response variable is y_i [112].
To estimate the coefficients β_j, the S function (i.e. the sum of squared residuals at the points in the data set) has to be minimized:

$$S = \sum_{i=1}^{N} \epsilon_i^2 \tag{3.61}$$

The residuals are the difference between the actual responses and the predicted ones at the locations x_i in the design space, and can be written as:

$$\epsilon_i = y_i - \hat{f}(x_i, \beta), \quad i = 1, \ldots, N \tag{3.62}$$

The minimum of S can be easily found by setting the gradient with respect to β equal to zero:

$$\frac{\partial S}{\partial \beta_j} = 2 \sum_{i=1}^{N} \epsilon_i \frac{\partial \epsilon_i}{\partial \beta_j} = -2 \sum_{i=1}^{N} \left[y_i - \hat{f}\left(x_i, \beta\right) \right] \frac{\hat{f}\left(x_i, \beta\right)}{\partial \beta_j} = 0, \quad j = 1, \ldots, m \tag{3.63}$$

Least square problems can be divided into two categories: linear and nonlinear [113, 111].

Linear problems

Linea problems have a closed-form solution, however they are not accurate and their main usage is for guessing the main trends of the response variable. Considering a problem with N experiments and k parameters, the metamodel function would have the form reported in equation 3.58, or the form of matrix notation:

$$y = X\beta + \epsilon \tag{3.64}$$

where:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{1,1} & \cdot & \cdot & x_{1,k} \\ 1 & x_{2,1} & \cdot & \cdot & x_{2,k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & x_{N,1} & \cdot & \cdot & x_{N,k} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \beta_N \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_N \end{bmatrix} \tag{3.65}$$

Equation 3.61 can then be written as:

$$S = \epsilon^t \epsilon = (y - X\beta)^T (y - X\beta) = y^T y - 1\beta^T X^T y + \beta X^T X\beta \tag{3.66}$$

Deriving equation 3.66 and equalling to zero yields to:

$$\frac{\partial S}{\partial \beta} = -2X^T y + 2X^T X\beta = 0 \tag{3.67}$$

Solving in β we obtain:

$$\beta = \left(X^T X\right)^{-1} X^T y \tag{3.68}$$

and the response of the estimated (fitted) metamodel is:

$$\hat{y} = X\beta \tag{3.69}$$

Nonlinear problems

Nonlinear problems should be solved iteratively. The initial values for the coefficients $\beta^{(1)}$ are chosen. Then the vector is iteratively updated (Gauss-Newton algorithm), obtaining at iteration k:

$$\beta^{(k+1)} = \beta^{(k)} + \Delta\beta^{(k)} \tag{3.70}$$

where $\Delta\beta^{(k)}$ is called shift vector. To update this vector we can use an iterative model by approximation to a first-order Taylor series expansion about $\beta^{(k)}$

$$\begin{aligned}\hat{f}\left(x_i,\beta^{(k+1)}\right) = \hat{f}\left(x_i,\beta^{(k)}\right) + \sum_{j=1}^{m}\frac{\partial\hat{f}\left(x_i,\beta^{(k)}\right)}{\partial\beta_j}\left(\beta_j^{(k+1)}-\beta_j^{(k)}\right) = \\ = \hat{f}\left(x_i,\beta^{(k)}\right) + \sum_{j=1}^{m} J_{i,j}^{(k)}\Delta\beta_j^{(k)}\end{aligned} \tag{3.71}$$

where **J** is the $N \times m$ Jacobian matrix of $\hat{f}$ with respect to β. This equation (3.71) can be written in the form of matrix notation as:

$$y = \hat{y}^{(k)} + \epsilon^{(k)} = \hat{y}^{(k+1)} + \epsilon^{(k+1)} = \hat{y}^{(k)} + J^{(k)}\Delta\beta^{(k)} + \epsilon^{(k+1)} \tag{3.72}$$

Subsequently, derivative of S with respect to β becomes:

$$\frac{\partial S}{\partial\beta} = -2J^{(k)^T}\epsilon^{(k)} + 2J^{(k)^T}J^{(k)}\Delta\beta^{(k)} \tag{3.73}$$

Solving in $\Delta\beta^{(k)}$ we obtain:

$$\Delta\beta^{(k)} = \left(J^{(k)^T}J^{(k)}\right)^{-1}J^{(k)^T}\epsilon^{(k)} \tag{3.74}$$

This method generally applies to any model function. However, for simplicity, its use is restricted to complete or incomplete polynomials. More complex and irregular functions may require the experimental evaluation of the Jacobian matrix, which can be achieved only with a huge amount of additional experimental work.
The quality of an approximating response surface can be estimated by regression parameters. The regression parameters are defined so that their values fall within the range $[0, 1]$ and the nearest they are to 1, the better the model is expected to be. A widely used regression parameter is the normal regression parameter R^2, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y_i,\hat{y}_i\right)^2}{\sum_{i=1}^{N}\left(y_i,\overline{y}\right)^2} \tag{3.75}$$

where

$$\overline{y} = \frac{\sum_{i=1}^{N} y_i}{N} \tag{3.76}$$

The adjusted regression parameter is defined as the normal regression parameter multiplied for a term depending on the DoE sample size and the number of coefficients m of the model function:

$$R^2_{adj} = 1 - \frac{\sum_{i=1}^{N} (y_i, \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i, \overline{y})^2} \frac{N-1}{N-m} \tag{3.77}$$

from which is clear that $R^2_{adj} \leq R^2 \leq 1$ and $\lim_{N\to\infty} R^2_{adj} = R^2$. To estimate the predictive capability of the model, N response surfaces are built in which one of the DOE sample points x_i is missing, then the prediction of the new response surface in x_i is evaluated by the predicted regression paramters:

$$R^2_{pred} = 1 - \frac{\sum_{i=1}^{N} (y_i, \hat{y}'_i)^2}{\sum_{i=1}^{N} (y_i, \overline{y})^2} \frac{N-1}{N-m} \tag{3.78}$$

where $\hat{y}'_i$ is the response of the model in which the sample point x_i is missing [93].

3.5.2 Optimal response surface methodology

Optimal Response Surface Methodology (O-RSM) is a generalization of the LSM [93]. Let us assume we want to build a least square response surface using the results of an experiment (x_i, y_i), with m basis functions X_j and m coefficients β_j, so that the sum of squared errors in

$$y = \hat{f}\left(x, \beta, X^{'}(x)\right) + \epsilon(x) = \sum_{j=1}^{m} \beta_j X_j(x) + \epsilon(x) \tag{3.79}$$

at the sample points x_i4,$i = 1, \ldots, N$ is minimized. In optimal rsm the optimal basis functions and their coefficients are to be determined. The optimal basis functions are chosen from a set $X(x) = \left[X_1(x), \ldots, X_p(x)\right]^T$, with $p > m$, where the terms $X_j(x)$ can be any function of x.
O-RSM is an iterative procedure in which, at iteration l the basis functions $X^{'(l)}(x)$ are randomly chosen from $X(x)$ and so the least squares response surface is computed:

$$\hat{y}^{(l)} = \hat{f}\left(x, \beta^{(l)}, X^{'(l)}(x)\right) = \sum_{j=1}^{m} \beta_j^{(l)} X_j^{(l)}(x) \tag{3.80}$$

A performance parameter r_i, $i = 1, \ldots, p$, is defined for each term in $X(x)$ and initially set to zero. After each iteration, the performance parameters of the basis function involved in the iteration are set to

$$r_i = r_i + \delta^{(l)} \tag{3.81}$$

where $\delta^{(l)}$ is a measure of the performance of the response surface at iteration l (i.e. any regression parameter).
With a large number of iterations, a heuristic estimation of the best basis functions is given by those elements in $X(x)$ whose performance parameter divided by the number of times the basis function has been chosen during the iterations is maximum. The O-RSM is given by the least squares function

$$\hat{y} = \hat{f}\left(x, \beta, X'^{best}(x)\right) = \sum_{j=1}^{m} \beta_j X_j^{best}(x) \tag{3.82}$$

where X'^{best} is the vector of the best basis functions.

3.5.3 Kriging

Kriging is a method of spatial interpolation developed by Lev Gandin in 1959 and named after South African mining engineer Daniel Gerhardus Krige [114]. Kriging is one of several methods that use a limited set of sampled data points to estimate the value of a variable over a continuous spatial field.
In Kriging method the estimation of the response variable at a point x is given by the linear combination of the results of a DoE run:

$$\hat{f}(x) = \sum_{i=1}^{N} \lambda_i(x) y_i \tag{3.83}$$

where λ_i is called weight for the measured value at i-th location. With the kriging method, the weights are based not only on the distance between the measured points and the prediction location but also on the overall spatial arrangement of the measured points. There are several sub-types of kriging according to the way μ, the expected value is computed, such as [115]:

- **simple Kriging** which assumes a known constant trend $\mu(x) = 0$
- **ordinary Kriging** which assumes an unknown constant trend $\mu(x) = \mu$
- **universal Kriging** which assumes a linear trend $\mu(x) = \sum_{j=1}^{k} \beta_j x_{i,j}$
- **block Kriging** which estimates averaged values over gridded "blocks" rather than single points. These blocks often have smaller prediction errors than are seen for individual points.
- **IRF-k Kriging** which assumes $\mu(x)$ ti be an unknown polynomial
- etc.

Ordinary Kriging is one of the most commonly used Kriging techniques and is described by the acronym BLUE, which stands for Best Linear Unbiased Estimator. It

is the "best" because it aims at minimising the variance of the errors, "linear" because its estimates are weighted linear combinations of the available data, and "unbiased" since it tries to have the mean residual or error equal to 0. Mathematically speaking, the variance is expressed as:

$$\begin{aligned}\hat{\sigma}^2 &= var\left(\hat{f}(x) - f(x)\right) = var\left(\hat{f}(x)\right) + var\left(f(x)\right) - 2 \cdot cov\left(\hat{f}(x), f(x)\right) \\ &= var\left(\sum_{i=1}^{N} \lambda_i(x) f(x_i)\right) + var\left(f(x)\right) - 2 \cdot cov\left(\sum_{i=1}^{N} \lambda_i(x) f(x_i), f(x)\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i(x)\lambda_j(x) cov\left(f(x_i), f(x_j)\right) + var\left(f(x)\right) \\ &\quad -2\sum_{i=1}^{N} \lambda_i(x) cov\left(f(x_i), f(x)\right) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i(x)\lambda_j(x) c\left(x_i, x_j\right) + var\left(f(x)\right) - 2\sum_{i=1}^{N} \lambda_i(x) c\left(x_i, x\right)\end{aligned} \tag{3.84}$$

where $c(x, y)$ is the covariance function. The variance is minimized under the unbiasedness condition, which in case of ordinary Kriging leads to:

$$\sum_{i=1}^{N} \lambda_i(x) = 1 \tag{3.85}$$

meaning that the sum of all weights is equal to 1.
Deriving equation 3.84 yields to:

$$c(x_i, x_j)\lambda(x) = c(x_i, x) \quad \rightarrow \quad \lambda(x) = c^{-1}(x_i, x_j) c(x_i, x) \tag{3.86}$$

from which, in order to determine $\lambda(x)$, $c(x_i, x_j)$ and $c(x_i, x)$ have to be estimated by means of a semivariogram model.
A semivariogram (γ) is a visual depiction of the covariance exhibited between each pair of points in the sampled data, and is used to determine spatial dependece. It can be computed via the following equation [116]:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i,j=1}^{N(h)} \left(z_i - z_j\right)^2 \tag{3.87}$$

where $N(h)$ stands for the number of pair observations (i, j) separated by a spatial distance h. The terms z_i and z_j are the attribute values of observation i and j respectively. This function calculates the attribute difference between neighbouring observations separated by a lag hh to evaluate if these observations display the same information. Regarding the semi-variogram, it means that as the distance between

observations increases, the semi-variance is likely to increase because near observations share more characteristics than distant ones. In figure 3.11, the black points sum up the spatial structure of the entire dataset. In fact, the semi-variogram is computed for all pairs of observations but the plot would be unreadable if all the semi-variances were reported. One of the strongest assumption of the variogram is that of second-order stationarity, the fact that the variogram is bounded and reaches a plateau [116, 117, 114].

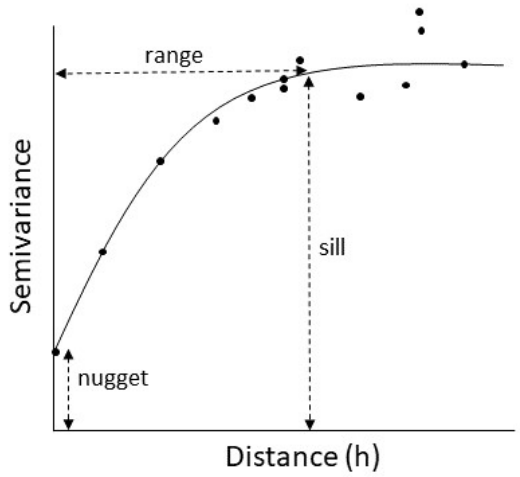


FIGURE 3.11: Theoretical exponential semi-variogram model

The **nugget effect**, often referred to as C_0, represents the small-scale spatial variations within the fields. This is an indicator of how noisy the spatial structure is. The **partial sill**, C_1, represents the magnitude of variation of the variable of interest. Intuitively, the higher the partial sill compared to the nugget, the stronger the spatial structure. The sill is the variance of the dataset and can be computed as the sum of the partial sill and nugget. The sill corresponds to the value when the semi-variances reaches a plateau and stabilizes. Last but not least, **the range**, a, is the distance beyond which observations are no longer spatially correlated. It is considered that in average, above a specific spatial distance and whatever the pair of points examined, observations are too dissimilar and do not share any relationship.
Several semi-variogram model exist, amongst which the most common are spherical, gaussian, circular, exponential and linear.
Having chosen the semi-variogram that best fit the analyzed data, it can then be used to determine $\lambda(x)$ by defining the covariances of equation 3.86 as:

$$c(x,y) = C_0 + C_1 - \gamma(h) \tag{3.88}$$

Including the unbiasedness condition for ordinary Kriging given by equation 3.85 into 3.86 leads to:

$$\begin{pmatrix} \gamma(x_1,x_1) & \cdot & \cdot & \gamma(x_1,x_N) & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \gamma(x_N,x_1) & \cdot & \cdot & \gamma(x_N,x_N) & 1 \\ 1 & \cdot & \cdot & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} c(x_1,x) \\ \cdot \\ c(x_N,x) \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1(x) \\ \cdot \\ \lambda_N(x) \\ \varphi \end{pmatrix} \tag{3.89}$$

that is the system which is usually solved in order to compute the weights vector λ at the new location, in which φ is the Lagrange multiplier.
Finally the value of $\hat{f}(x)$ is given by equation 3.83 and the prediction variance at x is [93]:

$$var\left(\hat{f}(x) - f(x)\right) = \begin{pmatrix} \lambda_1(x) \\ . \\ \lambda_N(x) \\ \varphi \end{pmatrix}^T \begin{pmatrix} \gamma(x_1, x) \\ . \\ \gamma(x_N, x) \\ 1 \end{pmatrix} = \lambda(x)^T \gamma(x_i, x) + \varphi \tag{3.90}$$

3.5.4 Concluding remarks

It is not easy to draw conclusions on RSM methods, as response surfaces are essentially interpolations or approximation of an unknown function. Since this function is not known, and the number of samples of the DoE are in general relatively low, the goodness of the response surface is not something that we are able to know precisely. Moreover, the choice of the control parameters impacts most of the methods, making the choice of the RSM method even more uncertaint.
However, when choosing the RSM there are some important things to consider:

- expected **noise** of the response variable: interpolation or approximation method. In most cases interpolation methods are preferable because if the response variable is not particularly noisy, at least in a certain neighbourhood of the DoE samples the estimation error is likely to be low. However, if we expect the noise to be significant, interpolating can be dangerous and can easily lead to unreliable responses. Amongst the methods described, LSM and O-RSM are approximating methods, while Kriging can be both interpolating or approximating depending on the nugget value.
- Expected **regularity** of the response variable. If something is known about the response variable, this could help in choosing the most appropriate method. For example, LSM would be a good choice if the response variable is expected to be polynomial. If the response is expected to be a fairly regular function then an interpolating method can be chosen. On the other hand, the choice of a proper RSM methodology for a very irregular function can be extremely hard, unless a large amount of data from the DoE is available.
- **Choice of the parameters**. This is a key aspect, since it has significant influence on the response surface itself. From this point of view, LSM is quite easy to treat since it only requires the definition of the order of the approximating polynomial. On the other hand, for Kriging the choice of the function and the parameters is not easy, but sistematic procedures exist to take over this issue. As a general rule, itmust be kept in mind that the most meaningful parameters, which can affect dramatically the output of the response surface, are those defining the distance within which the influence of a DOE sample is perceived.

- **Computational effort**. Building a RSM will require a computational effort, which in general is not an issue when compared to the time require for running the experiments or the simulation, however can become an important aspect. The computational effort grows with the DoE sample size, is almost null for LSM and a bit higher for O-RSM and Kriging.
- **Aim of the RSM**. Before choosing the RSM to use it is important to have clear in mind which is the aim of our study.

As a general statement, Kriging method always gives quite good response surfaces. If the response surface is expected to be quite regular also a LSM polynomial surface usually fits the data fairly enough. The additional complication of the LSM given by the O-RSM it is not worthy to be tried unless the shape of the response variable is likely to follow the shape of some of the functions chosen as a basis.
Since the computational effort needed for building response surfaces in general is not an issue, it is suggested to build up many surfaces using different methods and different sets of parameters, to compare them, and, if possible, to test their effectiveness versus a few more experimental results before choosing the one which seems to fit better.

References

[23] A. Cuneo, A. Traverso, and S. Shahpar. "Comparative analysis of methodologies for uncertainty propagation and quantification". In: *Proceedings of the ASME Turbo Expo*. Vol. 2C-2017. American Society of Mechanical Engineers, 2017. ISBN: 9780791850800. DOI: 10.1115/GT2017-63238. URL: https://asmedigitalcollection.asme.org/GT/proceedings/GT2017/50800/Charlotte,NorthCarolina,USA/241782.

[27] N. Metropolis and S. Ulam. "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247 (1949), pp. 335–341. ISSN: 1537274X. DOI: 10.1080/01621459.1949.10483310.

[28] N. Wiener. "The Homogeneous Chaos". In: *American Journal of Mathematics* 60.4 (1938), p. 897. ISSN: 00029327. DOI: 10.2307/2371268.

[30] M. D. McKay, R. J. Beckman, and W. J. Conover. "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 15372723. DOI: 10.1080/00401706.1979.10489755.

[34] D. Xiu and G. Em Karniadakis. "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos". In: *Computer Methods in Applied Mechanics and Engineering* 191.43 (2002), pp. 4927–4948. ISSN: 00457825. DOI: 10.1016/S0045-7825(02)00421-8.

[51] A. K. Gupta, T. F. Móri, and G. J. Székely. "How to transform correlated random variables into uncorrelated ones". In: *Applied Mathematics Letters* 13.6 (2000), pp. 31–33. ISSN: 08939659. DOI: 10.1016/S0893-9659(00)00050-1.

[52] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf. "Global sensitivity analysis for systems with independent and/or correlated inputs". In: *Journal of Physical Chemistry A* 114.19 (2010), pp. 6022–6032. ISSN: 10895639. DOI: 10.1021/jp9096919.

[53] M. Rosenblatt. "Remarks on a Multivariate Transformation". In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729394.

[54] H. Li and D. Zhang. "Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods". In: *Water Resources Research* 43.9 (2007). ISSN: 00431397. DOI: 10.1029/2006WR005673.

[55] J. L. Lumley. "The structure of inhomogeneous turbulence". In: *Atmospheric Turbulence and Radio Wave Propagation, edited by A. M. Yaglom and V. I. Tatarski (Nauka, Moscow)*. 1967, pp. 166–178. ISBN: 9783937655239.

[56] M. Eldred, C. Webster, and P. Constantine. "Evaluation of Non-Intrusive Approaches for Wiener-Askey Generalized Polynomial Chaos". In: 2008. DOI: 10.2514/6.2008-1892.

[57] K. Kim, M. R. Von Spakovsky, M. Wang, and D. J. Nelson. "Dynamic optimization under uncertainty of the synthesis/design and operation/control of a proton exchange membrane fuel cell system". In: *Journal of Power Sources* (2012). ISSN: 03787753. DOI: 10.1016/j.jpowsour.2011.11.014.

[58] D. Bigoni, A. P. Engsig-Karup, and H. True. "Comparison of classical and modern uncertainty qualification methods for the calculation of critical speeds in railway vehicle dynamics". In: *Proceedings of the Mini Conference on Vehicle System Dynamics, Identification and Anomalies*. Vol. 2012-Novem. 2012, pp. 91–100. ISBN: 9789633131022.

[59] K. P. Burnham and B. Efron. "The Jackknife, the Bootstrap and Other Resampling Plans." In: *Biometrics* 39.3 (1983), p. 816. ISSN: 0006341X. DOI: 10.2307/2531123.

[60] M. R. Chernick. *Bootstrap Methods*. 2007. DOI: 10.1002/9780470192573.

[61] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. 2009. ISBN: 9780387781648. arXiv: arXiv:1011.1669v3.

[62] J. H. Halton. "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals". In: *Numerische Mathematik* 2.1 (1960), pp. 84–90. ISSN: 0029599X. DOI: 10.1007/BF01386213.

[63] J. M. Hammersley. "Monte Carlo Methods for Solving Multivariable Problems". In: *Annals of the New York Academy of Sciences* 86.3 (1960), pp. 844–874. ISSN: 17496632. DOI: 10.1111/j.1749-6632.1960.tb42846.x.

[64] J Peter and M Marcelet. "Comparison of surrogate models for turbomachinery design". In: *WSEAS Transactions on Fluid Mechanics* 3.1 (2008), pp. 10–17. URL: http://dl.acm.org/citation.cfm?id=1353862.1353870.

[65] P. L'Ecuyer. "Quasi-monte Carlo methods in finance". In: *Proceedings - Winter Simulation Conference*. Vol. 2. 2004, pp. 1645–1655. DOI: 10.1109/wsc.2004.1371512.

[66] C. Joy, P. P. Boyle, and K. S. Tan. "Quasi-Monte Carlo methods in numerical finance". In: *Management Science* 42.6 (1996), pp. 926–938. ISSN: 00251909. DOI: 10.1287/mnsc.42.6.926.

[67] B. J. Collings and H. Niederreiter. "Random Number Generation and Quasi-Monte Carlo Methods." In: *Journal of the American Statistical Association* 88.422 (1993), p. 699. ISSN: 01621459. DOI: 10.2307/2290359.

[68] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. 2004. ISBN: 978-0-387-76369-9.

[69] W. J. Morokoff and R. E. Caflisch. "Quasi-Monte carlo integration". In: *Journal of Computational Physics* 122.2 (1995), pp. 218–230. ISSN: 00219991. DOI: 10.1006/jcph.1995.1209.

[70] M. B. Giles. "Multilevel Monte Carlo path simulation". In: *Operations Research* 56.3 (2008), pp. 607–617. ISSN: 0030364X. DOI: 10.1287/opre.1070.0496.

[71] S. Mishra, C. Schwab, and J. Šukys. "Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions". In: *Journal of Computational Physics* 231.8 (2012), pp. 3365–3388. ISSN: 10902716. DOI: 10.1016/j.jcp.2012.01.011.

[72] D. C. Montgomery. *Design and Analysis of Experiments Eighth Edition*. 2012. ISBN: 9781118146927. DOI: 10.1198/tech.2006.s372.

[73] L. Cassettari, P. G. Giribone, M. Mosca, and R. Mosca. "The stochastic analysis of investments in industrial plants by simulation models with control of experimental error: Theory and application to a real business case". In: *Applied Mathematical Sciences* 4.73-76 (2010), pp. 3823–3840. ISSN: 1312885X.

[74] L. Cassettari, R. Mosca, and R. Revetria. "Monte Carlo simulation models evolving in replicated runs: A methodology to choose the optimal experimental sample size". In: *Mathematical Problems in Engineering* 2012 (2012), pp. 73–76. ISSN: 1024123X. DOI: 10.1155/2012/463873.

[75] R. F. Gunst, R. H. Myers, and D. C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons, 2016. DOI: 10.2307/1270613.

[76] D. J. Pike, G. E. P. Box, and N. R. Draper. "Empirical Model-Building and Response Surfaces." In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 151.1 (1988), p. 223. ISSN: 09641998. DOI: 10.2307/2982196.

[77] L. Cassettari, R. Mosca, and R. Revetria. "Experimental error measurement in monte carlo simulation". In: *Handbook of Research on Discrete Event Simulation Environments: Technologies and Applications*. 2009, pp. 92–142. ISBN: 9781605667744. DOI: 10.4018/978-1-60566-774-4.ch006.

[78] U. Brussel. "Comparison of intrusive and non-intrusive polynomial chaos methods for CFD applications in aeronautics". In: *European Conference on Computational Fluid Dynamics* June (2010), pp. 14–17.

[79] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. 1991. DOI: 10.1007/978-1-4612-3094-6.

[80] O. P. Le Maitre and O. M. Knio. *Spectral methods for uncertainty quantification*. 2010. ISBN: 9789048135196.

[81] D. Xiu and G. Em Karniadakis. "The Wiener-Askey polynomial chaos for stochastic differential equations". In: *SIAM Journal on Scientific Computing* 24.2 (2003), pp. 619–644. ISSN: 10648275. DOI: 10.1137/S1064827501387826.

[82] H. N. Najm. *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*. 2009. DOI: 10.1146/annurev.fluid.010908.165248.

[83] M. S. Eldred and J. Burkardt. "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification". In: *47th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*. 2009. ISBN: 9781563479694. DOI: 10.2514/6.2009-976.

[84] I. Babuška, R. Temponet, and G. E. Zouraris. "Galerkin finite element approximations of stochastic elliptic partial differential equations". In: *SIAM Journal on Numerical Analysis* 42.2 (2004), pp. 800–825. ISSN: 00361429. DOI: 10.1137/S0036142902418680.

[85] S. Oladyshkin and W. Nowak. "Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion". In: *Reliability Engineering and System Safety* 106 (2012), pp. 179–190. ISSN: 09518320. DOI: 10.1016/j.ress.2012.05.002.

[86] H. Scheffe. *The analysis of variance*. John Wiley & Sons, 1999.

[87] G. D. Ruxton. "The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test". In: *Behavioral Ecology* 17.4 (2006), pp. 688–690. ISSN: 10452249. DOI: 10.1093/beheco/ark016.

[88] T. K. Kim. "T test as a parametric statistic". In: *Korean Journal of Anesthesiology* 68.6 (2015), pp. 540–546. ISSN: 20057563. DOI: 10.4097/kjae.2015.68.6.540.

[89] S. P. Schacht, J. E. Aspelmeier, S. p. Schacht, and J. E. Aspelmeier. "One-Way Analysis of Variance (ANOVA)". In: *Social and Behavioral Statistics*. 2018, pp. 195–214. DOI: 10.4324/9780429497308-11.

[90] J. W. Tukey. "Comparing Individual Means in the Analysis of Variance". In: *Biometrics* 5.2 (1949), p. 99. ISSN: 0006341X. DOI: 10.2307/3001913.

[91] *F critical value for ANOVA test*. URL: https://web.ma.utexas.edu/users/davis/375/popecol/tables/f005.html (visited on 01/08/2021).

[92] T. K. Kim. "Understanding one-way anova using conceptual figures". In: *Korean Journal of Anesthesiology* 70.1 (2017), pp. 22–26. ISSN: 20057563. DOI: 10.4097/kjae.2017.70.1.22.

[93] M. Cavazzuti. *Optimization methods: From theory to design scientific and technological aspects in mechanics*. 2013, pp. 1–262. ISBN: 9783642311871. DOI: 10.1007/978-3-642-31187-1.

[94] C. C. Craig and R. A. Fisher. "The Design of Experiments." In: *The American Mathematical Monthly* 43.3 (1936), p. 180. ISSN: 00029890. DOI: 10.2307/2300364.

[95] M. I. Rodrigues and A. F. Iemma. *Experimental design and process optimization*. 2014. ISBN: 9781482299564. DOI: 10.1201/b17848.

[96] S. L. Ferreira et al. "Box-Behnken design: An alternative for the optimization of analytical methods". In: *Analytica Chimica Acta* 597.2 (2007), pp. 179–186. ISSN: 00032670. DOI: 10.1016/j.aca.2007.07.011.

[97] G. Tagushi, Y. Yokoyama, and Y. Wu. *Taguchi Methods: Design of Experiments*. 1993.

[98] H. Evangelaras, C. Koukouvinos, and M. V. Koutras. "Advances in Robust Parameter Design: From Taguchi's Inner-Outer Arrays to Combined Arrays". In: *Encyclopedia of Statistical Sciences*. 2011. DOI: 10.1002/0471667196.ess7146.

[99] W. L. Loh. "On latin hypercube sampling". In: *Annals of Statistics* 24.5 (1996), pp. 2058–2080. ISSN: 00905364. DOI: 10.1214/aos/1069362310.

[100] J. C. Helton and F. J. Davis. "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems". In: *Reliability Engineering and System Safety* 81.1 (2003), pp. 23–69. ISSN: 09518320. DOI: 10.1016/S0951-8320(03)00058-9.

[101] M. Stein. "Large sample properties of simulations using latin hypercube sampling". In: *Technometrics* 29.2 (1987), pp. 143–151. ISSN: 15372723. DOI: 10.1080/00401706.1987.10488205.

[102] A. Olsson, G. Sandberg, and O. Dahlblom. "On Latin hypercube sampling for structural reliability analysis". In: *Structural Safety* 25.1 (2003), pp. 47–68. ISSN: 01674730. DOI: 10.1016/S0167-4730(02)00039-5.

[103] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. 1997. DOI: 10.1137/1.9780898719574.

[104] G. W. S. and D. S. Watkins. "Fundamentals of Matrix Computations." In: *Mathematics of Computation* 59.199 (1992), p. 299. ISSN: 00255718. DOI: 10.2307/2153000.

[105] A. M. Law and W. D. Kelton. *Simulation modeling and analysis*. 1991. ISBN: 0780379241. DOI: 10.1016/j.sysconle.2007.02.002. arXiv: 0608329v1 [arXiv:astro-ph].

[106] N. Bradley. "Response Surface Methodology". PhD thesis. 2007. ISBN: 9780444527011. DOI: 10.1016/B978-044452701-1.00083-1.

[107] J. P. Kleijnen and R. G. Sargent. "A methodology for fitting and validating metamodels in simulation". In: *European Journal of Operational Research* 120.1 (2000), pp. 14–29. ISSN: 03772217. DOI: 10.1016/S0377-2217(98)00392-0.

[108] M. Mäkelä. "Experimental design and response surface methodology in energy applications: A tutorial review". In: *Energy Conversion and Management* 151 (2017), pp. 630–640. ISSN: 01968904. DOI: 10.1016/j.enconman.2017.09.021.

[109] J. P. Kleijnen. "An overview of the design and analysis of simulation experiments for sensitivity analysis". In: *European Journal of Operational Research* 164.2 (2005), pp. 287–300. ISSN: 03772217. DOI: 10.1016/j.ejor.2004.02.005.

[110] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira. "Response surface methodology (RSM) as a tool for optimization in analytical chemistry". In: *Talanta* 76.5 (2008), pp. 965–977. ISSN: 00399140. DOI: 10.1016/j.talanta.2008.05.019.

[111] J. C. F. Gauss. *Combinationis observationum erroribus minimis obnoxiae*. University of Gottingen, 1825.

[112] M. A. Hariri-Ardebili, S. M. Seyed-Kolbadi, and M. Noori. "Response Surface Method for Material Uncertainty Quantification of Infrastructures". In: *Shock and Vibration* 2018 (2018). ISSN: 10709622. DOI: 10.1155/2018/1784203. URL: https://doi.org/10.1155/2018/1784203.

[113] E. H. Lockwood and A. L. Edwards. "An Introduction to Linear Regression and Correlation". In: *The Mathematical Gazette* 69.447 (1985), p. 62. ISSN: 00255572. DOI: 10.2307/3616472.

[114] N. Cressie. "The origins of kriging". In: *Mathematical Geology* 22.3 (1990), pp. 239–252. ISSN: 08828121. DOI: 10.1007/BF00889887.

[115] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[116] G. Bohling. *Introduction to geostatistics and variogram analysis*. October. 2005.

[117] N. Cressie. "Fitting variogram models by weighted least squares". In: *Journal of the International Association for Mathematical Geology* 17.5 (1985), pp. 563–586. ISSN: 00205958. DOI: 10.1007/BF01032109.

Chapter 4

Analysis of uncertainties in compact plate-fin recuperators for microturbines

4.1 Objective

The aim of this first study [40] was to perform a stochastic analysis on microturbine compact recuperators to evaluate the impact of uncertainties in design parameters on their cost and volume, and to fill the literature gap between a traditional deterministic approach to recuperator design and a probabilistic approach, applied to compact recuperators for mGTs.
Two different probabilistic modelling approaches (Monte Carlo and Response Sensitivity Analysis) have been developed within Matlab® and applied to evaluate the impact of pressure drops and effectiveness uncertainties on relevant recuperator features, such as its volume and cost. As a matter of fact, a preliminary study of the manufacturing uncertainties has already been carried out by HiFlux [118], a company which designs and manufactures compact heat exchangers. This allowed to quantify the uncertainties on recuperator key parameters, such as pressure drop and effectiveness, which are due to tolerances of parts and manufacturing process. This work focuses on the outcome of the aforementioned industrial data to evaluate how such uncertainties impact on cost and volume of a case-study recuperator.
Indeed, Monte Carlo and Response Sensitivity Analysis have been coupled with Compact Heat Exchanger Optimisation and Performance Evaluation (CHEOPE) tool, published in [119] and outlined hereby, which allows to analyse two different kinds of recuperator: the furnace-brazed plate-fin type and the welded primary surface type. Such recuperator technologies are the most suitable for microturbine applications being extremely compact, with a ratio of heat transfer surface to heat exchanger volume greater than 700 m^2/m^3 [120].
In the first section of this chapter a brief introduction to mGT and recupators is carried out. Then the CHEOPE tool is described to give the reader an understanding of its operation, and finally the deterministic and stochastic analisys are presented alongside with the results obtained.

4.2 General Introduction to Recuperators

Micro-Gas Turbine (MGT) are a promising power generation technology for distributed generation and smart micro-grids [121, 122, 123, 124], as they present several advantages compared to piston engines in terms of low emissions, compact size, good reliability, fuel flexibility and low maintenance costs. However, their high capital cost is still hindering their deployment at large scale. A typical MGT system mainly consists of a cetrifugal compressor, a radial turbine, a combustor, a recuperator and a high speed generator, as can be observed in the following figure 4.1:

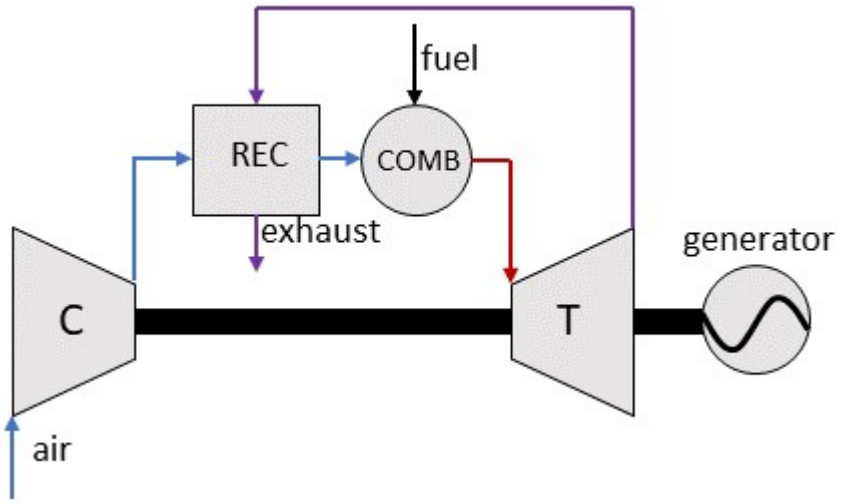


FIGURE 4.1: Micro-Gas Turbine system layout

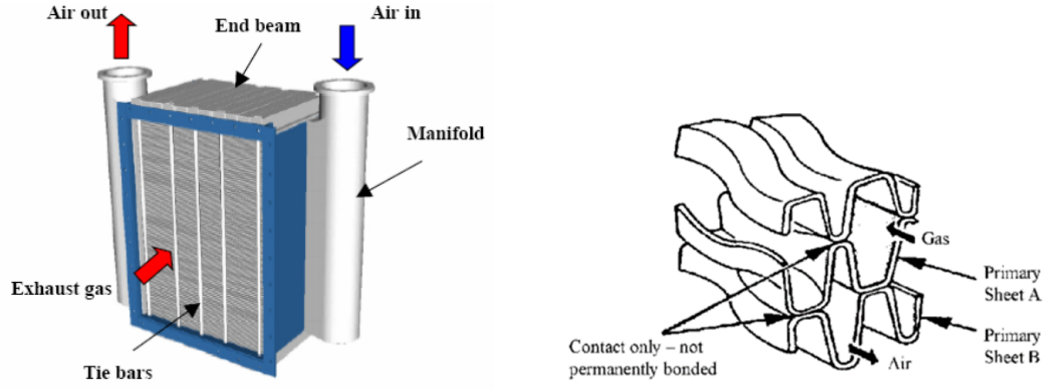
To achieve an electrical efficiency of at least 30% the use of a gas-gas heat exchanger, or recuperator, is mandatory [125], and so the proper design of this component becomes a key point. As a matter of fact, such a recuperator must present several important technical characteristics such as high effectiveness, compact size, structural integrity, and minimal pressure drops. Such requirements cannot be considered free-of-use and indeed the recuperator can represent up to 30% of the total capital cost of a microturbine package [125]. The proper design of the recuperator becomes a central point in microturbine design as well as the evaluation of its performance, as it can be widely inferred from open literature [126, 127, 128].
One of the central points in recuperator performance studies is the identification of the key parameters affecting its volume and, ultimately, cost. Several authors, such as Stevens et al. [129], and Lagerström and Xie [130] recognised effectiveness and pressure drops as key parameters. Despite the huge research efforts in this field, most of the analyses on recuperators are performed without considering the uncertainties which affect some of the variables on which the recuperator design is based, since they cannot be pre-determined with appropriate precision. These uncertainties can have a strong impact on the design of such a component, affecting both costs and performances [23], making the quantification and the evaluation of their impact on the outputs of interest a key aspect of the design.

4.3 Recuperator model and CHEOPE

Recuperator performance is calculated with the CHEOPE tool, originally developed by the Thermochemical Power Group (TPG) from the University of Genoa (UNIGE), which allows one to perform two types of analyses on gas-gas heat exchangers:

- Rating: off-design analysis
- Sizing: design and cost estimation

Rating analysis allows one to evaluate the recuperator performance in off-design conditions, computing the effectiveness and the heat exchanger temperatures while changing the mass flows and the inlet temperatures. Sizing analysis allows one to estimate the main geometrical features of the heat exchanger based on design operating conditions (effectiveness, pressure drops, gas mass flows, inlet gas temperatures and heat transfer surface characteristics), providing the properties of the gas outlets, heat exchange volume and estimated cost.



(A) plate-fin surface [130] (B) cross-wavy surface [131]

FIGURE 4.2: Recuperator configurations available in CHEOPE

Regarding the plate-fin configuration figure 4.2 A, the first step done by CHEOPE is calculation of the required plate thickness to resist the pressure differential between the two sides. A detailed mechanical and thermal stress analysis, along with creep and fatigue, would normally be involved in this calculation. Given the scope of this work, the following simplified approach is followed. The plate is considered as an equivalent beam, fixed at both ends, with the same thickness t and length l equal to the fin pitch p_f, which is subject to a uniformly distributed load due to the pressure difference Δp_{side}. The equivalent section resistance modulus W is expressed by eq. 4.1, while eq. 4.2 gives the resultant thickness [132].

$$W = \frac{lt^2}{6} \tag{4.1}$$

$$t = p_f \sqrt{\frac{\Delta p_{side}}{2\tau_d}} \tag{4.2}$$

The design mechanical stress (τ_d) can be set based on a known material for plate-fin recuperator. A similar approach is followed for the primary surface recuperator. The sizing procedure is based on the number of heat transfer units [133] for both recuperator types. To calculate the heat-transfer convective coefficients h and the fanning factors f the following equations have been used. In particular, for the plate-fin recuperator (figure 4.2 A), the equations were derived from [134]:

$$h_{plate-fin} = 0.6522 \cdot G \cdot c_p \cdot Pr^{-\frac{2}{3}} \cdot Re^{-0.5403} \cdot \left(\frac{p_f - \delta}{b - \delta}\right)^{-0.1541} \left(\frac{\delta}{l}\right)^{0.1499} \left(\frac{\delta}{p_f - \delta}\right)^{-0.0678} \tag{4.3}$$

$$f_{plate-fin} = 9.6243 \cdot Re^{-0.7422} \cdot \left(\frac{p_f - \delta}{b - \delta}\right)^{-0.1856} \left(\frac{\delta}{l}\right)^{0.3053} \left(\frac{\delta}{p_f - \delta}\right)^{-0.2659} \tag{4.4}$$

and are expressed as a function of specific flow rate G, Prandtl number Pr, Reynolds number Re, specific heat c_p, fin thickness δ, distance between plates b and fin pitch p_f.
When considering primary surface recuperators instead, the equations were derived from [131]:

$$Nu_{ps} = k_1 + k_2 \cdot Re \tag{4.5}$$

$$f_{ps} = \frac{k_3}{Re} + k_4 \tag{4.6}$$

where the coefficients k_1, k_2, k_3, k_4 used depend on the type of surface geometry employed.
Leakages have been neglected in this analysis, as in a brand-new recuperator they should be negligible, and may occur only because of sub-component failure or thermal stress.
CHEOPE capability for plate-fins recuperator sizing was tested and validated in a previous work [119] through data available in open literature [134]. As underlined in the introduction, the recuperator cost cannot be considered negligible as it represents about 30% of the capital cost of a microturbine package [135]. Generally, detailed cost information is kept confidential by the recuperator industry, however, it is possible to develop a general cost equation that can provide a valid indication of the cost trend as different recuperator configurations are considered. The first step to establish a proper cost function is to highlight the main parameters influencing the recuperator capital cost and then gathering them in a proper equation for cost estimation.
It is reasonable to consider that the component cost depends directly on the quantity of material used for its realisation and, especially for plate-fin recuperators, inversely on the hydraulic diameter of the heat transfer surface employed. In fact, the lower the value of the hydraulic diameter (D_h), the higher is the degree of complexity for

the assembly, considering welding and brazing processes. So, the introduction of the hydraulic diameter in the cost function allows one to consider both the technology level and the complexity of the heat exchanger.
These considerations led to the following equation for cost estimation:

$$cost = \psi \left(\frac{D_{h,0}}{D_h} \right)^a \left(\frac{\rho Vol}{\rho_0 Vol_0} \right) \tag{4.7}$$

Where ψ represents the cost of the reference recuperator (labeled "0"), whose matrix weight is known ($\rho_0 Vol_0$). The exponent a, also called scaling factor, assumes the value of 0.6. The material volume is evaluated as the product of the heat transfer total surface (A) and the average thickness t, based on equation 8.

$$Vol = A \cdot t \tag{4.8}$$

Such an equation implies dependence on the heat transfer effectiveness, which affects the required surface (and thus the material volume) in a strongly non-linear way, especially for high effectiveness values ($> 85\%$).
The cost equation proposed has the following limitations of applicability: it can only be employed within the operating range of the material used in the reference recuperator and, in principle, it cannot be extended to recuperator concepts different from the reference one. However, within the range of validity, eq. 4.7 clearly states the proportionality between the cost and the quantity of material employed. The cost function has been validated and tuned properly in previous works [119] and updated to current state-of-the-art market [118].

4.4 Deterministic analysis

CHEOPE has been used to size a recuperator with the deterministic on-design parameters enlisted in Table 4.1, which are representative of a mGT cycle application [136].

TABLE 4.1: Recuperator on-design input data and characteristics

		Hot Side	**Cold Side**
Mass flow	$[kg/s]$	0.0719	0.0676
Inlet temperature	$[°C]$	593.6	186.2
Inlet pressure	$[bar]$	1.04	3.18
Pressure drop	$[\%]$	1.77	0.61
Effectiveness	$[-]$	0.8835	
Thermal conductivity	$[W/mK]$	20	
Metal plate thickness	$[mm]$	0.15	
Type of recuperator		plate-fin	
Surface type		S11.94D	
Flow direction		counter-flow	

The choice of a plate-fin recuperator is motivated by its high compactness and well established manufacturing process, featuring good design flexibility [120]. Moreover, such recuperators have already been used for a variety of mGT applications [125]. A strip-fin surface type with double structure (S11.94D) [133] has been assumed for both the hot and cold sides. To have consistent baseline data to validate the stochastic analysis, a simulation in deterministic conditions was performed with CHEOPE, using data listed in Table 4.1 and monitoring cost and volume as outputs.

TABLE 4.2: Recuperator determinstc outputs of interest

Outputs		
Cost	[€]	1978
Volume	$[dm^3]$	5.98

4.5 Analysis under uncertainties of the reference case

The stochastic analysis was performed assuming as stochastic inputs the pressure losses on the cold and hot sides and the effectiveness, considering their on-design value as mean and the standard deviation based on data provided by HiFlux [118]. The level of uncertainty for both inputs and outputs can be quantified with a Coefficient Of Variance (COV), defined as the ratio between standard deviation and the absolute value of the mean value (σ/μ). A high COV value denotes a large uncertainty; since it relates the standard deviation to the actual mean value, it can give more information than the absolute σ value.

TABLE 4.3: Mean and standard deviation of the stochastic inputs

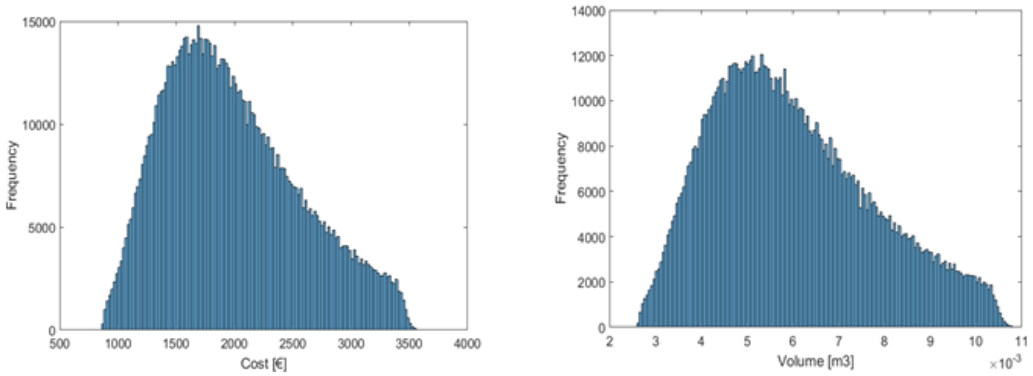
		μ	σ	COV
Pressure drop hot side	[%]	1.77	0.10	5.65
Pressure drop cold side	[%]	0.61	0.05	8.20
Effectiveness	$[-]$	0.8365	0.05	5.99

A normal distribution was assumed for all the uncertain variables considered. For most engineering problems, a clear probability inference of parameters usually requires a large volume of experimental data, which is often impractical due to expense consideration or experimental limitations. Thus, a normal (Gaussian) distribution is popularly adopted without losing the generality, proving to be, under such circumstances, more appropriate than other distributions.

4.5.1 Monte Carlo analysis

A Monte Carlo simulation was performed with one million samplings. The values of the input variables (pressure drop on the cold and hot sides and effectiveness) are extracted from normal distributions with the features presented in Table 3. The number of samplings was chosen to ensure the smallest Mean Square Pure Error

(MSPE) possible (Section 3.2.1). Cost and volume of the recuperator were monitored as outputs and the following probability density functions (Figure 4.3), means and standard deviations (Table 4) were obtained.



(A) Cost PDF ($\mu = 1982$€) (B) Volume PDF ($\mu = 5.993 dm^3$)

FIGURE 4.3: Probability density functions of the recuperator outputs

TABLE 4.4: MCs outputs

		μ	σ
Cost	[€]	1982	581
Volume	$[dm^3]$	5.993	1.758

The probability density functions of cost and volume resemble the shape of a Weibull distribution and it can be observed that they both present a wide range of variability, suggesting a high dependence on input uncertainties. However, the results obtained from MC, in terms of mean, are in line with the deterministic outputs calculated previously (4.2), and by increasing the number of runs the results did not change significantly, confirming the fact that the number of runs performed were adequate and that MC is a valid and accurate method to perform analysis of the model under uncertainty. However, this method requires a huge amount of computational time, making it unsuitable for applications that require both simulation accuracy and low computational efforts. To overcome this issue, the RSA method was applied to the model, as explained in the following section.

4.5.2 Response Sensitivity Analysis results

The Response Sensitivity Analysis confirmed the substantial advantage in terms of time to obtain the results. In fact, the RSA method bases its calculation algorithm on knowledge of the characteristics of the main probabilistic input variables (mean and standard deviation) and therefore does not require any sampling. Such inherent simplicity means that the time required to perform the analysis depends only on the number of variables influencing the output of the model. In this study the output variables (cost and volume) were estimated applying a second order central finite difference scheme for the mean (Section 3.2.2) and a first order central finite difference scheme for the variance (Section 3.2.2). To tune the RSA method properly, a percentage error on the mean was defined as:

$$\%err_{\mu_i} = \frac{\mu_i - \mu_{MC}}{\mu_{MC}} \cdot 100 \tag{4.9}$$

Where the reference mean is the one computed by MC (Table 4.2). The same equation was applied to the standard deviation. Table 4.5 results were obtained with ten simulation runs, i.e. several orders of magnitude lower than MC.

TABLE 4.5: RSA mean and standard deviation (error is estimated against MC results)

		μ	σ	$\%err_\mu$	$\%err_\sigma$
Cost	[€]	1936	631	−2.36	7.8
Volume	$[dm^3]$	5.77	1.91	−3.86	7.8

These results are in line with those obtained with MC, both on cost and volume, as the percentage errors on means and standard deviations, as defined in eq. 4.9, are below 10%, indicating that the RSA tuning was done properly and so this stochastic method can be used to evaluate the impact of uncertainties on the recuperator analysed.
By evaluating the sensitivity (Section 3.2.2), it can be observed that the cost and volume are highly dependent on the effectiveness (Figure 4.4), while the pressure drop on the hot side has very little impact on them and the pressure drop on the cold side has no influence on the monitored outputs. The sensitivity sign indicates whether the relationship between the output and the input considered is direct or inverse. Effectiveness has then a direct impact on cost and volume, while pressure drop on the hot side is inversely related to them, as expected (i.e. a decrease would lead to higher cost and volume).

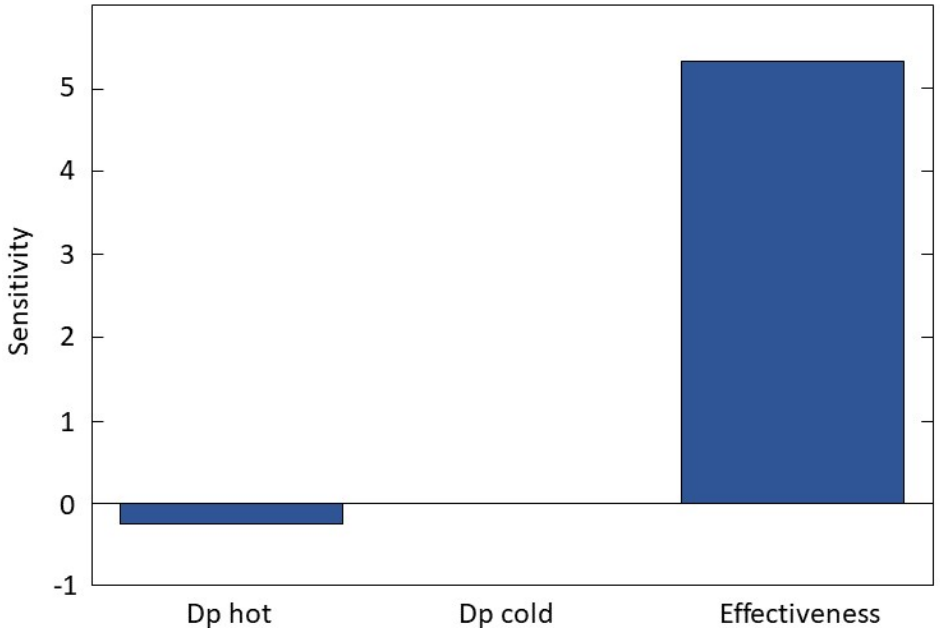


FIGURE 4.4: Recuperator sensitivity from RSA layout

Analysing the resulting sensitivity, and the variance of the output in RSA, it can

be stated that the monitored outputs are influenced almost exclusively by the uncertainty in effectiveness. Therefore, it was decided to assess the impact of decreasing the effectiveness standard deviation by a certain percentage. In such a case, Table 4.6 demonstrates that the standard deviation of the cost decreases almost linearly with the effectiveness one (i.e. a 10% decrease of the effectiveness standard deviation leads to a 10% decrease in the standard deviation of the cost). On the other hand, the mean presents a slight increase due to the second term of the equation with which it is computed (Section 3.2.2).

TABLE 4.6: Impact of decreasing effectiveness standard deviation on recuperator cost

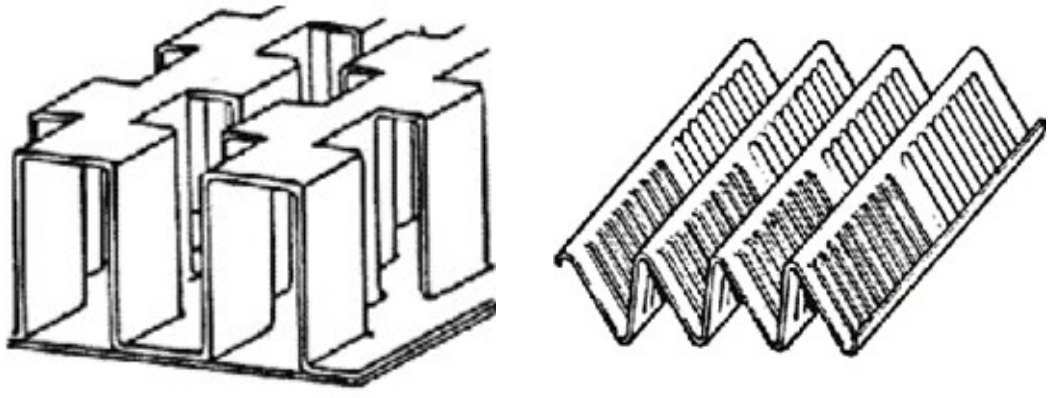
Case study	σ_ϵ	μ_{cost} [€]	$\%diff.$	σ_{cost} [€]	$\%diff.$
Reference case	0.050	1936	–	631	–
$\sigma_\epsilon - 10\%\sigma_\epsilon$	0.045	1945	0.4	568	−10
$\sigma_\epsilon - 20\%\sigma_\epsilon$	0.040	1952	0.8	505	−20

4.6 Analysis of cost and volume of CHEOPE recuperators under uncertainty

The RSA method was then applied to all the plate-fin recuperators, using all the surfaces available in the CHEOPE embedded library, which presented a metal plate thickness analogue to the reference recuperator (Table 1). The mean and standard deviation of costs and volumes were computed for each recuperator surface type, indicated in Table 7 with a reference code where the first letter refers to the fin type:

- S: Offset Strip Fin (figure 4.5A)
- L: Louvered Fin (figure 4.5B)
- P: Plain Fin (figure 4.2A)

and the last letter identifies the kind of structure used (D: Double Structure, T: Triple Structure) [133].



(A) Offset strip fin recuperator (B) Louvered fin recuperator

FIGURE 4.5: Probability density functions of the recuperator outputs

The parameter β_A represents the area density defined as the ratio between the heat transfer surface and the heat exchanger volume. In the following tables (Tables 4.7-4.9) and figures (Figures 4.6-4.7) the solutions which present the lowest mean costs are identified with yellow, while the recuperators with the lowest mean volumes are identified with green colour. All the recuperators analyzed in the following have been evaluated using the on-design conditions listed in Table 4.1, for what concerns mass flows, inlet temperatures and inlet pressures.

TABLE 4.7: RSA results of CHEOPE recuperators

Surface type	D_h $[mm]$	β_A $[m^2/m^3]$	μ_{cost} [€]	σ_{cost} [€]	μ_{volume} $[dm^3]$	σ_{volume} $[dm^3]$
S11.1	3.08	1204	1696	554	7.81	2.56
S11.94D	2.27	1512	1936	631	5.77	1.91
S15.2	2.65	1368	2282	721	8.14	2.55
S15.4D	1.61	2106	1861	606	3.21	1.06
S16.00D	1.86	1804	1869	602	4.05	1.32
S16.12D	1.55	2165	1767	574	2.89	0.95
S16.12T	1.57	2133	2107	673	3.75	1.14
P5.3	6.15	617	2567	829	31.55	10.21
P11.1	3.08	1204	2095	686	9.63	3.17
P14.77	2.59	1378	2472	794	8.49	2.78
P15.08	2.67	1358	3054	970	10.89	3.47
P19.86	1.87	1841	2335	756	4.93	1.63
P11.94T	2.86	1289	2268	742	9.42	3.04
P12.00T	2.87	1288	2131	696	8.88	2.87
P16.96T	1.72	1994	2609	844	4.79	1.59
P25.79T	1.15	2807	2427	787	2.51	0.82
P30.33T	1.22	2666	3176	999	3.59	1.12
L6.06	4.45	840	1902	628	14.72	4.80
L8.70	3.65	1007	1653	539	7.63	2.50
L11.1	3.08	120	1696	554	7.97	2.57

The percentage COV for the recuperators analysed was evaluated (Figure 4.6), both for the cost and for the volume. Between the different configurations, the mean value of the COV is 32.36 for the cost and 32.42 for the volume, with a maximum variability of ± 0.88 and ± 2.10 respectively. Hence, the COV between the different configurations can be approximated as constant. However, comparing these values with that of input, it is possible to note that the COV increases by about one order of magnitude in respect to the COV related to the pressure drop at the cold side (the higher one in the inputs, see Table 4.1). This increase means that the actual design tends to increase the input uncertainties through the outputs. Analysing the recuperators in a Volume-Cost chart (Figure 4.7) is it possible to highlight that the recuperator which employs surface $S16.12D$ on both the cold and hot side, represents the best compromise between low cost and low volume in this first analysis. As CHEOPE allows one to size a recuperator with different surfaces on the hot and cold sides, two further analyses were carried out:

1. Combination of surfaces from the three minimum-cost recuperators (Table 4.8)
2. Combination of surfaces from the three minimum-volume recuperators (Table 4.9)

In both analyses the cost and the volume were monitored as outputs.

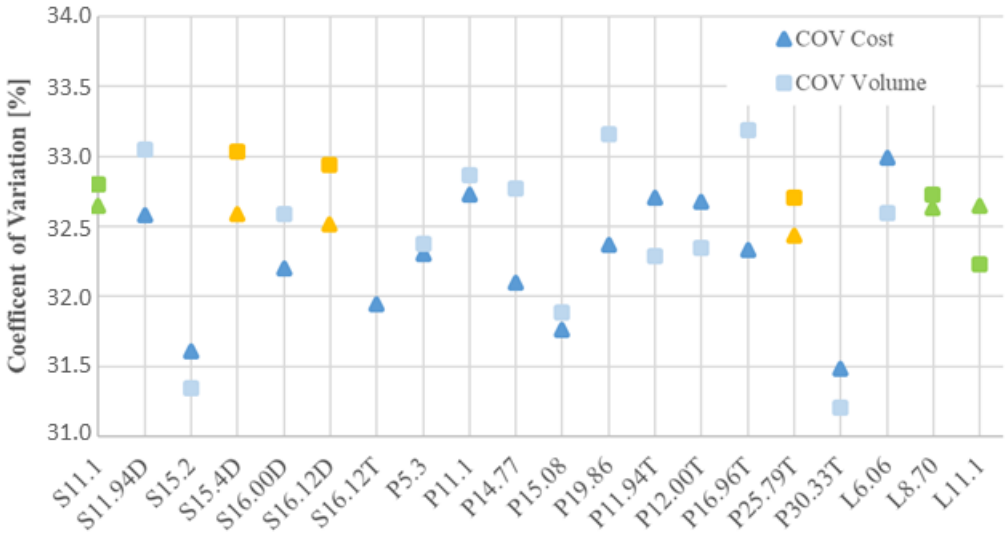


FIGURE 4.6: COV of analyzed recuperators

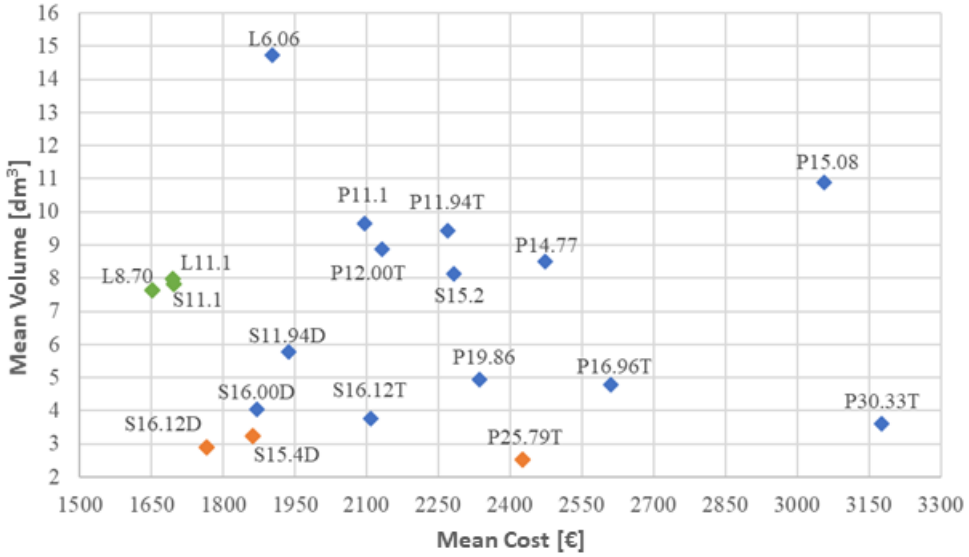


FIGURE 4.7: Volume and cost of each analyzed recuperator

From Table 4.8, it is possible to highlight that the solution with the lowest cost is the one which employs surface $L8.70$ on both sides (cold and hot), leading to a mean cost of €1652.60. This solution, among the low-cost recuperators, is also the one with the lowest volume ($7.63dm^3$). However, all the combinations exploited in Table 4.8 present small differences in terms of cost and volume between them, with a maximum variation of €43.2 and $0.34dm^3$ respectively. On the other hand, from Table 4.9 it can be observed that the solution which presents the minimum volume is the one which employs surface $P25.79T$ on both sides (cold and hot), with a volume of $2.52dm^3$. However, this solution is also the most expensive among the combinations investigated: in fact, it has the smallest hydraulic diameter ($1.15mm$), impacting the cost function described in eq. 4.7. The solution, among the minimum-volume combinations, with the lowest cost is the one which employs surface $S15.4D$ on the hot side and surface $S16.12D$ on the cold side. In the following figure (Figure 4.8) a final

TABLE 4.8: RSA - minimum cost recuperators

Hot side surface	**Cold side surface**	μ_{cost} [€]	σ_{cost} [€]	μ_{volume} $[dm^3]$	σ_{volume} $[dm^3]$
S11.1	S11.1	1696.25	533.68	7.81	2.56
S11.1	L8.70	1659.05	541.39	7.65	2.51
S11.1	L11.1	1696.75	553.84	7.76	2.56
L8.70	S11.1	1690.36	551.69	7.80	2.55
L8.70	L8.70	1652.60	539.22	7.63	2.50
L8.70	L11.1	1690.88	551.87	7.96	2.56
L11.1	S11.1	1695.30	553.37	7.82	2.56
L11.1	L8.70	1658.15	541.09	7.65	2.51
L11.1	L11.1	1695.80	553.54	7.97	2.57

TABLE 4.9: RSA - minimum volume recuperators

Hot side surface	**Cold side surface**	μ_{cost} [€]	σ_{cost} [€]	μ_{volume} $[dm^3]$	σ_{volume} $[dm^3]$
S15.4D	S15.4D	1861.33	606.47	3.21	1.06
S15.4D	S16.12D	1758.14	572.23	2.89	0.97
S15.4D	P25.79T	2107.57	686.84	2.71	0.95
S16.12D	S15.4D	1881.68	612.58	3.24	1.05
S16.12D	S16.12D	1766.63	574.38	2.89	0.95
S16.12D	P25.79T	2113.65	688.15	2.80	0.92
P25.79T	S15.4D	2175.39	705.68	2.85	0.93
P25.79T	S16.12D	2040.43	660.94	2.77	0.86
P25.79T	P25.79T	2426.81	787.08	2.52	0.82

comparison between the most valuable recuperators, according to the previous analyses, was done. Two further recuperators were evaluated, combining the surface of the minimum-cost recuperator with that of minimum-volume both on the cold and hot sides. From the results obtained (Figure 4.8) it can be observed that the solution which employs surface $P25.79T$ on the hot side and surface L8.70 on the cold side presents extremely high costs compared to the others and also a high volume, while the solution which employs surface $L8.70$ on the hot side and surface P25.79T on the cold side is not particularly interesting as it presents intermediate characteristics between recuperator $P25.79T - P25.79T$ and $L8.70 - L8.70$ but with a higher cost and volume than recuperator type $S16.12D$. As a matter of fact, the solution with the recuperator surface type $S16.12D$ on both the cold and hot sides continues to be that which presents the best compromise between cost and volume. This solution, compared to that with the lowest volume ($P25.79T - P25.79T$) presents a cost of about 27% lower and a volume 15% higher, while compared to the lowest-cost solution ($L8.70 - L8.70$) has a volume 13% lower and a cost 37% higher. The solution which employs surface $L8.70$ on the hot side and surface P25.79T on the cold side is not particularly interesting as it presents intermediate characteristics between recuperator $P25.79T - P25.79T$ and $L8.70 - L8.70$ but with a higher cost and higher volume than recuperator type $S16.12D$.

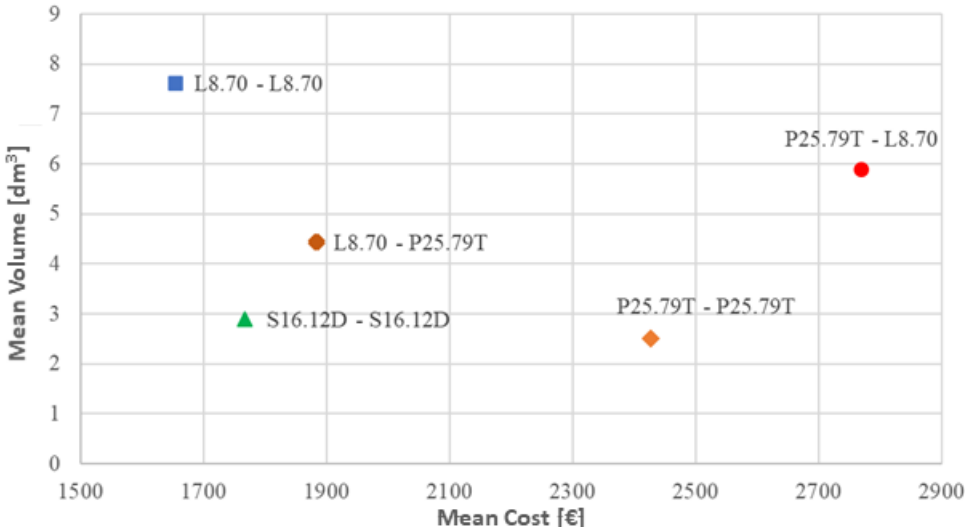


FIGURE 4.8: Comparison of the most valuable recuperators in terms of cost and volumeS

4.7 Concluding remarks

In this first work, developed during the first year of the Ph.D. study, a stochastic analysis of a microturbine compact recuperator has been performed. Two different probabilistic methods have been applied and coupled with CHEOPE, a design tool developed by the TPG-UNIGE to size compact recuperators, to evaluate the impact of uncertainties in input parameters on cost and volume of a case-study plate-fin recuperator. Three main parameters have been considered as uncertain: effectiveness, hot side pressure drop and cold side pressure drop, based on industrial evidence. Monte Carlo simulation has allowed the author to evaluate the probability density function of the monitored outputs and has highlighted a high variability of cost and volume. The application of the RSA method has allowed to further investigate the root cause, highlighting the effectiveness as the main parameter affecting the results, while the hot side pressure drop presented a small influence and the cold side pressure drop no relevant impact at all. In fact, it was assessed that decreasing the uncertainty on the effectiveness standard deviation by 10 to 20% would decrease the standard deviation of the cost by about the same amount. Further RSA analyses were performed on the recuperators within the inner library of CHEOPE, considering those with the same metal thickness and the same design points as the case study plate-fin recuperator analysed previously, to identify the least expensive and the most compact solution, combining different surfaces on the cold and hot sides. The COV of the recuperators analysed was almost constant, however its increase of about one order of magnitude against the stochastic inputs COV, highlighted that the actual design amplifies the input uncertainties through the outputs. In particular the minimum-volume recuperator, which employs surface $P25.79T$ on both the cold and hot sides, has a mean volume of $2.52dm^3$, while the mean cost is €2486.81. On the other hand, the least expensive recuperator employs surface L8.70 on both sides with a mean cost of €1652.60, but with a volume of $7.63dm^3$. In the last analysis a comparison between the most compact solutions and the least expensive has been done, identifying recuperator $S16.12D$ as the best compromise between cost and volume. This solution presents a volume of $2.89dm^3$ and a cost of €1766.63,

which, in comparison with the minimum volume and the minimum cost configurations, is 15% higher in volume and 37% higher in cost, respectively. Finally, this paper showed the potential of coupling a detailed sizing tool such as CHEOPE with uncertainty quantification methods, which can now be effectively used for the robust design of compact gas-gas recuperators.

References

[23] A. Cuneo, A. Traverso, and S. Shahpar. "Comparative analysis of methodologies for uncertainty propagation and quantification". In: *Proceedings of the ASME Turbo Expo*. Vol. 2C-2017. American Society of Mechanical Engineers, 2017. ISBN: 9780791850800. DOI: 10.1115/GT2017-63238. URL: https://asmedigitalcollection.asme.org/GT/proceedings/GT2017/50800/Charlotte,NorthCarolina,USA/241782.

[40] A. Giugno, A. Cuneo, and A. Traverso. "Analysis of uncertainties in compact plate-fin recuperators for microturbines". In: *Applied Thermal Engineering* 150.September 2018 (2019), pp. 1243–1251. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2019.01.093. URL: https://doi.org/10.1016/j.applthermaleng.2019.01.093.

[118] Hiflux. *Hiflux recuperator*. URL: https://hiflux.co.uk/applications/microturbine-recuperators/.

[119] A. Traverso and A. F. Massardo. "Optimal design of compact recuperators for microturbine application". In: *Applied Thermal Engineering* 25.14-15 (2005), pp. 2054–2071. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2005.01.015.

[120] R. Shah. "Compact heat exchangers for microturbines". In: *Micro Gas Turbines* Educationa.September (2005), pp. 1–18. URL: http://services.bepress.com/cgi/viewcontent.cgi?article=1026{\&}context=eci/heatexchangerfall2005.

[121] G. Xiao, T. Yang, H. Liu, D. Ni, M. L. Ferrari, M. Li, Z. Luo, K. Cen, and M. Ni. *Recuperators for micro gas turbines: A review*. 2017. DOI: 10.1016/j.apenergy.2017.03.095.

[122] M. L. Ferrari, M. Pascenti, L. Magistri, and A. F. Massardo. "Micro gas turbine recuperator: Steady-state and transient experimental investigation". In: *Proceedings of the ASME Turbo Expo*. 2009. ISBN: 9780791848869. DOI: 10.1115/GT2009-59172.

[123] L. Galanti and A. F. Massardo. "Micro gas turbine thermodynamic and economic analysis up to 500 kWe size". In: *Applied Energy* 88.12 (2011), pp. 4795–4802. ISSN: 03062619. DOI: 10.1016/j.apenergy.2011.06.022.

[124] W. De Paepe, D. Coppitters, S. Abraham, P. Tsirikoglou, G. Ghorbaniasl, and F. Contino. "Robust Operational Optimization of a Typical micro Gas Turbine". In: *Energy Procedia*. Vol. 158. 2019, pp. 5795–5803. DOI: 10.1016/j.egypro.2019.01.549.

[125] C. F. McDonald. "Recuperator considerations for future higher efficiency microturbines". In: *Applied Thermal Engineering* 23.12 (2003), pp. 1463–1487. ISSN: 13594311. DOI: 10.1016/S1359-4311(03)00083-8.

[126] M. L. Ferrari, A. Sorce, M. Pascenti, and A. F. Massardo. "Recuperator dynamic performance: Experimental investigation with a microgas turbine test rig". In: *Applied Energy* 88.12 (2011), pp. 5090–5096. ISSN: 03062619. DOI: 10.1016/j.apenergy.2011.07.016.

[127] P. Maghsoudi, S. Sadeghi, H. Khanjarpanah, and H. H. Gorgani. "A comprehensive thermo-economic analysis, optimization and ranking of different microturbine plate-fin recuperators designs employing similar and dissimilar fins on hot and cold sides with NSGA-II algorithm and DEA model". In: *Applied Thermal Engineering* 130 (2018), pp. 1090–1104. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2017.11.087.

[128] B. J. Tsai and Y. L. Wang. "A novel Swiss-Roll recuperator for the microturbine engine". In: *Applied Thermal Engineering* 29.2-3 (2009), pp. 216–223. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2008.02.028.

[129] T. Stevens, F. Verplatesen, and M. Baelmans. "Requirements for recuperators in micro gas turbines". In: *PowerMEMS Conference*. 2004, pp. 96–99.

[130] G. Lagerström and M. Xie. "High performance & cost effective recuperator for micro-gas turbines". In: *American Society of Mechanical Engineers, International Gas Turbine Institute, Turbo Expo (Publication) IGTI*. 2002. DOI: 10.1115/gt2002-30402.

[131] E. Utriainen and B. Sundén. "A comparison of some heat transfer surfaces for small gas turbine recuperators". In: *Proceedings of the ASME Turbo Expo*. 2001. ISBN: 9780791878521. DOI: 10.1115/2001-GT-0474.

[132] E. Avallone, T. Baumeister, and A. Sadegh. *Marks' standard handbook for mechanical engineers*. 2007. DOI: 10.5860/CHOICE.44-6870. URL: http://choicereviews.org/review/10.5860/CHOICE.44-6870.

[133] W. M. Kays and A. L. London. *Compact heat exchangers. Third Edition.* 1984. ISBN: 0070334188.

[134] R. M. Manglik and A. E. Bergles. "Heat transfer and pressure drop correlations for the rectangular offset strip fin compact heat exchanger". In: *Experimental Thermal and Fluid Science* 10.2 (1995), pp. 171–180. ISSN: 08941777. DOI: 10.1016/0894-1777(94)00096-Q.

[135] C. F. McDonald. "Low-cost compact primary surface recuperator concept for microturbines". In: *Applied Thermal Engineering* 20.5 (2000), pp. 471–497. ISSN: 13594311. DOI: 10.1016/S1359-4311(99)00033-2.

[136] M. L. Ferrari, M. De Campo, and L. Magistri. "Design and emulation of a turbocharged bio-fuelled sofc plant". In: *Proceedings of the ASME Turbo Expo*. 2018. ISBN: 9780791851043. DOI: 10.1115/GT201875026.

Chapter 5

Design and optimization under uncertainty of solid oxide fuel cell hybrid systems

5.1 Objective

The growing interest and sensibility towards the environment have impacted the world of research in the energy field, which has shifted focus towards conversion technologies, which could achieve higher efficiencies and guarantee a lower environmental impact. In such context, fuel cells are considered a promising technology for energy conversion, as they can guarantee high electrical efficiencies, low emissions, possibility for cogeneration and modularity [137, 138, 139]. In particular, Solid Oxide Fuel Cell (SOFC) are the most attractive fuel cell technology for stationary power production and hybrid system integration, thanks to their fuel flexibility and their high operating temperature, which can raise the overall cycle efficiency to more than 70%, resulting in a particularly interesting application for hybrid system integration [140, 141]. On the other side, in order to evaluate properly such systems and to get insights on their real behaviour, the range of performance and the economic profitability and investment which would be expected, a classic deterministic design would not be adequate. As a matter of fact, it is widely demonstrated that the performance of such systems is highly affected by uncertainties related to mechanical and manufacturing parameters, limited knowledge of physics and numerical approximations introduced with models [18, 23]. For these reasons, a single point deterministic simulation would mostly lead to an inaccurate or incomplete representation of the system considered. The objective of this chapter is to propose and assess a comprehensive and all-around framework for uncertainty analysis, robust design and multi-objective optimization of an hybrid system, considering different market conditions, which could affect the costs and the revenues, and making use of the response surface methodology to create a reliable metamodel on which perform these analyses. The first section describes briefly the layout of the hybrid system considered, while the second section describes more in detail the model employed, focusing on the main equations that drives the system in relation to the analyses which have been carried out.

5.2 Hybrid System layout

The system considered in this work is made up of a pressurized SOFC coupled with a turbocharger. In comparison with Micro-Gas Turbine (MGT) based hybrid system this solution provides less power, but it still benefits from cell pressurization: in fact, it is possible to increase a fuel cell efficiency by 11% raising its pressure from 1 to 5 bar [142]. A turbocharged SOFC features considerably lower capital cost, which is key to make hybrid systems commercially competitive in the future, thanks to the turbocharger technological maturity versus MGT [141].
The turbocharged SOFC system, which is presented in Figure 5.1, is powered by biogas with a 50% CH_4 - 50% CO_2 molar composition and it has been designed to provide a net electrical power output of about 30 kW. In this study a SOFC stack composed by 1500 cells is considered.
The fuel is compressed, pre-heated by the system exhausts in the Fuel Pre-Heater (FPH) and then lead to the anodic ejector primary nozzle. The purpose of the ejector is to recirculate the anode outlet to the reformer, in order to provide the proper amount of steam and heat for the chemical reactions and thus to avoid carbon deposition inside the SOFC. Inside the reformer, through Steam-Methane Reforming (SMR) (eq. 5.1) and Water-Gas Shift (WGS) reactions (eq. 5.2) [143], H_2 is obtained from biogas and it is used to feed the SOFC.

$$CH_4 + H_2O \rightarrow CO + 3H_2 \tag{5.1}$$

$$CO + H_2O \leftrightarrow CO_2 + H_2 \tag{5.2}$$

Cathodic and anodic flows from the SOFC outlet are mixed within the Off Gas Burner (OGB), where the fuel which has not been consumed inside the fuel cell is burnt. The heat produced by this combustion is provided both to pre-heat the cathodic inlet flow in the Air Pre-Heater (APH) and to drive the SMR chemical reaction in the reformer. Then the flow goes into the turbine and its expansion provides the mechanical power needed by the turbocharger to compress the air flow.
Before entering the SOFC, the air flow is pre-heated through the Recuperator (REC) and the APH. The reformer and all the heat exchangers are set up in a co-flow configuration.
In nominal operating conditions U_f is equal to 0.8, the fuel cell average temperature is $1071K$ and the anodic ejector recycle ratio (ratio between secondary and primary mass flow) is about 7.15. A high percentage (98.6%) of the methane in the anode flow reacts in the external reformer, while the remaining part is reformed inside the SOFC.

The system is equipped with a Waste-Gate Valve (WGV), which is used to lead part of the flow downstream from the turbine, without expansion. In this way, the flow at the REC inlet has a higher energy content and the air flow entering the fuel cell has a higher temperature. The opening of this valve is controlled to comply with SOFC operational constraints: the fuel cell maximum temperature must be between 1123 and $1133K$, while the stack internal temperature gradient must be lower than $250K$ [144]. The percentage of mass flow to bypass through the wastegate valve is

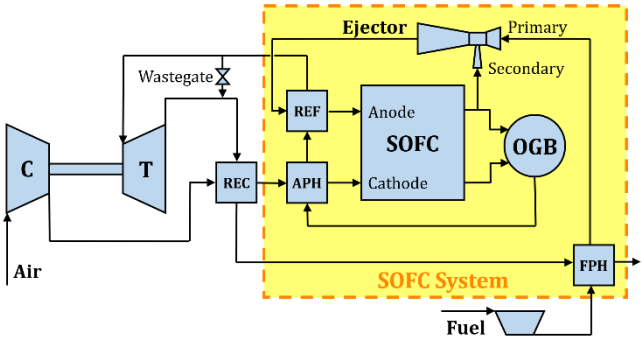


FIGURE 5.1: Turbocharged SOFC system plant layout

automatically determined by a Proportional Integral (PI) controller in order to keep the SOFC max temperature equal to 1133K [145].

5.3 Model Description

To analyze the hybrid system behavior under different design choices, a steady-state model was created in Matlab-Simulink®. All the components models are part of a library developed by TPG along the years and they are based on mass and energy balance equations. Inlet and outlet flows of each component are defined by mass flow, pressure, temperature and chemical composition.
Their validation was carried out in previous works [146, 147, 148, 149]. In [149] the validation of SOFC and Reformer (REF) was performed comparing simulations against experimental data collected onto a Staxera GmbH system. Validation of the SOFC model under faulty conditions was presented and discussed in [148]. The reliability verification of the heat exchanger model was carried out in [146], while the ejector was validated in [147].
The SOFC is modelled through a 1D approach, which allows to observe how temperatures and pressures vary along the cell, and thus to verify if all the operational limits are satisfied. Each cell in the stack is discretized in 10 finite elements and the electrochemical processes occurring in the SOFC are simulated as it follows:

- The value of current, set by the user, is used to compute the amount of H_2 and O_2 that are consumed
- Cathode and anode outlet compositions are obtained from the balances of the electrochemical reactions, neglecting the contributions of CO and CH_4
- The temperatures of the cell and of the cathode and anode flows are computed solving energy balance equations that consider the stack thermal losses
- The cell voltage is computed subtracting ohmic, mass transfer and activation losses from the Nernst's potential value.

The ohmic losses are directly related to the cell current i_{cell} and R_{dis}, which is computed according to eq. 5.3 for each discretization element. In this equation the

parameter k_{ohm} was introduced during the model validation to better match the experimental test data, while t and c are respectively the thickness and the conductivity of the part considered (anode, cathode and electrolyte) and A_{dis} represents the area of the discretized element.

$$R_{dis} = (\frac{t_{an}}{c_{an}} + \frac{t_{cat}}{c_{cat}} + \frac{t_{elec}}{c_{elec}}) \cdot \frac{k_{ohm}}{A_{dis}} \tag{5.3}$$

where k_{ohm} is a corrective coefficient that has been introduced to have simulations results as close as possible to the experimental behavior of the SOFC. In the following analyses the uncertainty regarding the ohmic losses magnitude is considered as related to k_{ohm}. The overall value of k_{ohm} is then calculated as the average between all the discretization elements.
The heat exchanger model is used to simulate APH, FPH and REC, but only the latter is a real heat exchanger, while APH and FPH represent heat exchanges between close pipes. This model is based on a 1D approach and the outlet flows temperatures are determined solving conductive and convective heat exchange equations.
The REF model is based on the same approach of the heat exchanger regarding the thermal exchange. The outlet flow composition is computed as a function of the heat provided and assuming equilibrium conditions for SMR and WGS reactions.
In the 0D ejector model the secondary nozzle pressure and the inlet flows are used to solve continuity, momentum and energy equations. In this way it is possible to determine primary nozzle inlet pressure and outlet stream properties. The OGB is simulated with a 0D model which computes outlet flow temperature and composition solving the fuel combustion reaction.
The fuel compressor is based on a 0D approach to find the pressurized fuel properties. Under the hypothesis of isentropic compression and perfect gas, the power absorbed by the fuel compressor $P_{C,fuel}$ is computed as:

$$P_{C,fuel} = \dot{m}_{fuel} \cdot (c_{p,out} T_{out} - c_{p,in} T_{in}) \tag{5.4}$$

While the electrical power generated by the SOFC is computed as:

$$P_{SOFC} = \eta_{inv} \cdot V_{real} \cdot i_{cell} \cdot N_{cell} \tag{5.5}$$

Subtracting $P_{C,fuel}$ from P_{SOFC} it is possible to obtain the net power output of the whole system P_{net}:

$$P_{net} = P_{SOFC} - P_{C,fuel} \tag{5.6}$$

while its η_{net} is computed according to eq. 5.7.

$$\eta_{net} = \frac{P_{net}}{(\dot{n}_{fuel} \cdot LHV_{fuel})} \tag{5.7}$$

Since V_{real} (eq. 5.5) is calculated as the difference between Nernst's potential and activation, ohmic and mass transfer losses, it is clear from eq. 5.5, eq. 5.6 and eq. 5.7 that ohmic losses have a direct influence on the system power output and efficiency. The steam to carbon ratio (STCR) at the anode ejector outlet, which must be maintained above 1.8 to avoid carbon deposition [150], is computed according to:

$$STCR = \frac{\dot{n}_{H_2O}}{\dot{n}_{CH_4} + \dot{n}_{CO}} \tag{5.8}$$

The turbocharger model is based on compressor and turbine 0D models. Both of them find values of efficiency, pressure ratio, mass flow and rotational speed interpolating or extrapolating component maps, which are based on real components maps, properly scaled to comply with the SOFC hybrid system requirements [148]. Since the turbocharger is not devoted to energy production, its power balance must always be zero (taking into account the mechanical losses). Shaft speed is obtained iteratively to ensure that this condition is met.
Finally, to better understand the analyses that will follow, it is important to underline that:

- the turbocharger is fixed, in terms of compressor and expander nominal capacity and related characteristic maps
- the fuel mass flow is determined based on the fuel cell area and the electrical current density.

In Table 5.1 it is possible to observe the most significant data and results from on-design simulation of the system described here under deterministic conditions (i.e. no uncertainty considered).

TABLE 5.1: Main input data and results from the on-design steady-state simulation of the hybrid system.

Input Data					
$i_{cell}[A]$	$\eta_t[-]$	$\eta_c[-]$	$\eta_{alt}[-]$	$k_{ohm}[-]$	$N_{cell}[-]$
30.30	0.6080	0.7168	0.960	0.480	1500

Results					
P_{net}	$[kW]$	30.2	$\dot{m}_{fuel}$	$[kg/s]$	0.0044
η_{net}	$[-]$	0.5198	$\dot{m}_{air}$	$[kg/s]$	0.0713
$T_{cell,max}$	$[K]$	1133	N_{shaft}	$[rpm]$	267853
$T_{cell,avg}$	$[K]$	1071	RR_{ej}	$[-]$	7.15
V_{real}	$[V]$	0.7148	k_{ohm}	$[-]$	0.48

5.4 Propagation of uncertainties

The first study conducted on the hybrids was performed in order to study the propagation of uncertainties in hybrid systems related both to the fuel cell (ohmic losses, anodic ejector diameter and gas composition) and the gas turbine cycle characteristics (compressor and turbine nominal efficiencies, recuperator pressure losses). Due to plant complexity and high computational effort required by uncertainty quantification methodologies, a response surface is created. In this work, a second-order design called Central Composite Design Face Centered (CCF) method with Least Square Method (LSM) is used (section 3.4.2 and 3.5.1). The choice of the LSM was done based on the fact that the response were expected to be regular, and the LSM would have guaranteed good results with an easy implementation. This design expands the idea of the first-order 2^k factorial design, where each control variable k is measured at two levels, called low and high levels, by adding center (n_0) and axial points for the creation of the metamodel. For the CCF the axial points are arranged so that they are at the center of each face of the factorial space. Thus, the total number of design points in a CCF is:

$$n_{runs} = 2^k + 2k + 1 \tag{5.9}$$

The impact of the following seven parameters, or factors:

- recuperator pressure drop at hot (dp_{hot}) and cold side (dp_{cold})
- anode ejector diameter (d_{ej})
- fuel cell ohmic losses corrective coefficient (k_{ohm})
- gas composition (CO_2 percentage)
- compressor (η_c) and turbine efficiency (η_t).

on five different system outputs, or responses:

- net power produced by the system (P_{net})
- net efficiency (η_{net})
- mass flow percentage bypassed through the WG valve ($\dot{m}_{WG}$)
- steam to carbon ratio (STCR)
- turbocharger rotational speed (N_{shaft})

is investigated. The creation of the RS for the UQ analysis is based on the knowledge related to the uncertainties of such input, based on industrial data, literature review and Authors' knowledge (table 5.3).

The RS was then created considering low and high levels to have a range equal to $\pm 3\sigma$ (table 5.3). Such range was considered since in a normal distribution, the interval created by the mean plus or minus three times the standard deviations contains 99.73% of the data.

A second-order response surface metamodel was built with a three-level factorial design, since a significant two-way interaction effect is observed at the end of

TABLE 5.2: Mean and coefficient of variance for the stochastic inputs - UQ analysis on hybrid system

Variable	Unit of measure	μ	COV
dp_{cold}	$[mbar]$	15	5% [151]
dp_{hot}	$[mbar]$	15	5% [151]
d_{ej}	$[m]$	0.0025	2.5%
k_{ohm}	$[-]$	0.48	3% [152]
CO_2	$[\%]$	50	5% [153]
η_c	$[\%]$	72	1%
η_t	$[\%]$	61	1%

TABLE 5.3: Low and high levels considered for RS creation - UQ analysis on hybrid system

Factor	Name	Central point	Low level	High level
A	dp_{cold}	15.00	12.75	17.25
B	dp_{hot}	15.00	12.75	17.25
C	d_{ej}	0.0025	0.0023	0.0027
D	k_{ohm}	0.4800	0.4368	0.5232
E	CO_2	0.5000	0.4250	0.5750
F	η_c	72.00	69.84	74.16
G	η_t	61.00	59.17	62.83

the ANOVA analysis. A total of 143 runs with the simulation model were then necessary for the RS creation, as required to perform a proper CCF (equation 5.9) where the number of factors k considered in this study is equal to 7. These runs were performed with the Matlab model described before.
Hereafter in table 5.4, the results related to the procedure needed to define the coefficient of the response surface is reported only for the net power, to give the reader an understanding of the method used.

Firstly, the ANOVA was performed to identify the most suitable model (i.e. 2F1, quadratic, cubic) and then to evaluate which single factors and which combinations were significant. In particular, the quadratic model, for all the response surfaces, was chosen for this study based on ANOVA results (table 5.4). As a recall, the generic equation for quadratic model is reported in the following:

$$y = \sum_{i,j=1}^{k} a_i x_i + b_i x_i x_j \quad \forall i, j \in \mathbb{N} \tag{5.10}$$

The significance of the quadratic model is based on p-values, R^2 and F-value. It can be observed that the model has a p-value lower than 0.0001, indicating the significance of the quadratic model. The absence of any lack-of-fit (p-value>0.05) in the other factors strengthens the reliability of the model. The Model F-value of 21764.16 implies the model is significant. The Predicted R^2 of 0.9996 is in reasonable agreement with the Adjusted R^2 of 0.9997. The Predicted R^2 represents the degree to which the input variables explain the variation of the output/predicted variable,

TABLE 5.4: ANOVA results for net power response surface - UQ analysis on hybrid system

Sources	Sum of squares	Mean square	F-value	p-value	
Model	60.77	3.20	21764.16	< 0.0001	significant
A	0.0007	0.0007	4.68	0.0325	
B	0.0066	0.0066	44.73	< 0.0001	
C	0.2830	0.2830	1926.02	< 0.0001	
D	24.18	24.18	1.645E05	< 0.0001	
E	34.99	34.99	2.381E05	< 0.0001	
F	0.3985	0.3985	2711.69	< 0.0001	
G	0.5789	0.5789	3939.59	< 0.0001	
CD	0.0113	0.0113	76.58	< 0.0001	
CE	0.0434	0.0434	295.04	< 0.0001	
CF	0.0175	0.0175	119.33	< 0.0001	
CG	0.0054	0.0054	36.47	< 0.0001	
DE	0.0347	0.0347	235.97	< 0.0001	
DF	0.0013	0.0013	8.55	0.0041	
DG	0.0062	0.0062	42.41	< 0.0001	
EF	0.0653	0.0653	444.45	< 0.0001	
EG	0.0116	0.0116	79.08	< 0.0001	
B^2	0.0010	0.0010	6.99	0.0093	
E^2	0.0304	0.0304	206.87	< 0.0001	
F^2	0.0018	0.0018	12.54	0.0006	
Residual	0.0181	0.0001			

while Adjusted R^2 gives the percentage of variation explained by only those independent variables that in reality affect the dependent variable. Typically, the more non-significant variables are added into the model, more the gap in R^2 and Adjusted R^2 increases. In table 5.5 the values of Adjusted R^2 for each response surfaces are reported. The high values indicate that all the models fit the simulation data very well.

TABLE 5.5: Adjusted R^2 values - UQ analysis on hybrid system

Response	R^2_{adj}
P_{net}	0.9997
η_{net}	0.9997
$\dot{m}_{WG}$	0.9996
STCR	0.9996
N_{shaft}	0.9974

5.4.1 Diagnostic analysis

Diagnostic plots are shown in Figure 5.24 to investigate the suitability of fit of the proposed response surface in comparison to the simulation model behaviour. Figure 5.24a represents the normal probability plot of the residuals. Since the points are detected on a straight line the normality of response data is proved [49]. Predicted

(values evaluated by the model) vs. actual values (values evaluated by the simulation model) plot proves that the predicted data are in agreement with observed data in the range of the operating variables (Figure 5.24b). Externally studentized residuals versus predicted values are shown in Figure 5.24c. Since the points are sited within a range close to zero, it means that an absence of constant error occurs. Figure 5.24d explored residual vs. run plot and shows random and uniform scatter of coloured points corresponding to P_{net}. The diagnostic plots of the other responses are reported in Annex A at the end of this section, however the trend is the same as the one observed for the net power.

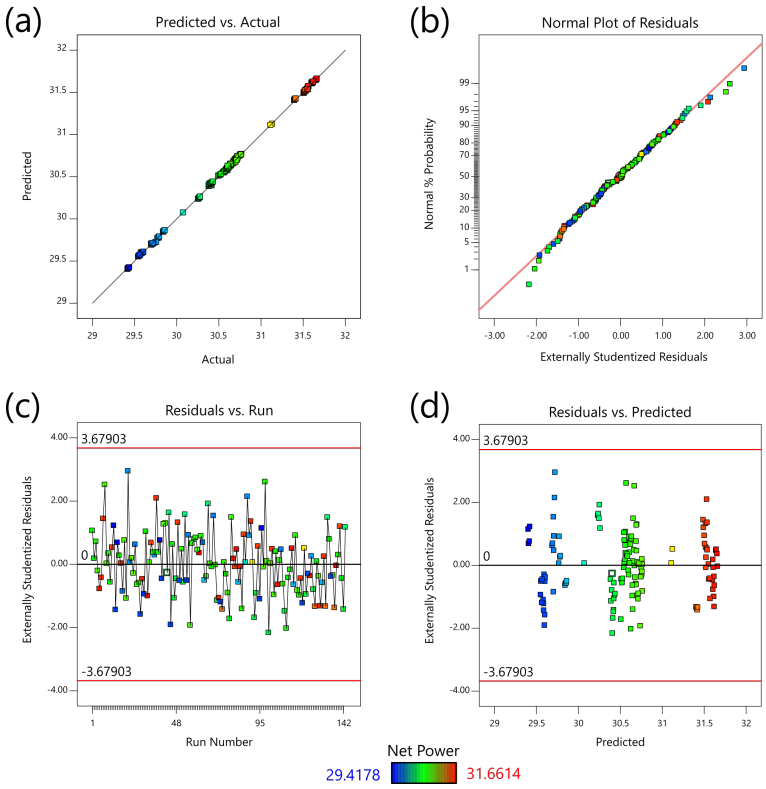


FIGURE 5.2: Diagnostic plots related to net power - UQ analysis on hybrid system

The obtained models can be used for predicting novel observations within the defined design range. Extrapolation should be avoided, as prediction variance increases dramatically.

5.4.2 Analysis of the responses

Having established the quality of the response surfaces, the same were used to study the effect of the analysed inputs on several performance parameters. Firstly, an analysis of the response surfaces is presented to indicate within which range the outputs can vary due to the uncertainties and how the response surfaces look like. Then, the interactions between the inputs and the main effects that inputs have on the responses are presented. This is the core of the RS methodology. In fact, such a quantitative evaluation of the effects and their interaction cannot be easily obtained with a sensitivity analysis or a Monte Carlo simulation. Finally, in the last section,

the results of a MCS applied to the RS are presented, to evaluate the probability distribution of the outputs as well as their COV. In the following response surfaces figures, it is worth to remind that the red points represent the points used for the response surface creation based on simulation model results.

Net power and net efficiency

Considering the uncertainties, net power varies between 29.4kW and 31.7kW (nominal is 30.71kW) with a net efficiency in the range from 49.8% to 53.6% (nominal is 51.98%). Looking at the ANOVA results (table 5.4) and Annex A, it is noticeable that these two response variables are closely correlated and their behaviour in relation to the input parameters is the same: for this reason, in the following just the figures related to the net power will be shown. As expected, the factors which have the strongest impact on them are the CO2 percentage in the fuel and the ohmic losses in the fuel cell, as they directly impact the performance of the SOFC stack (figure 5.3).

The P_{net} decreases along with the CO_2 increase are due both to a decrease in the

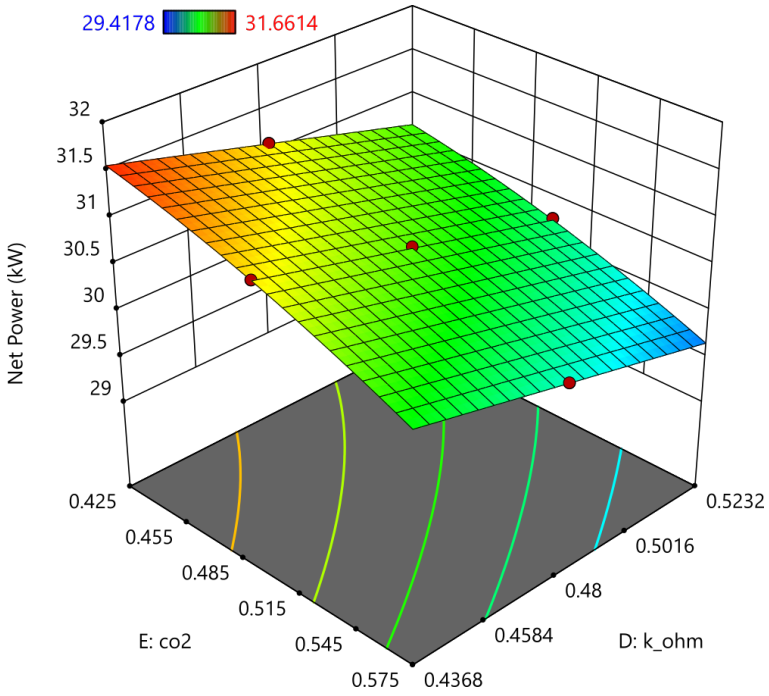


FIGURE 5.3: Net power response surface - UQ analysis on hybrid system)

SOFC power because of a lower amount of H_2, as well as to an increase in the power absorbed by the fuel compressor, as shown in figure 5.4.

Wastegate mass flow

The percentage of the nominal mass flow that passes through the WG valve varies between 2.5% and 18% (nominal is 11.5%). This means that, considering the uncertainties and the considerations made in the previous sections, the wastegate valve

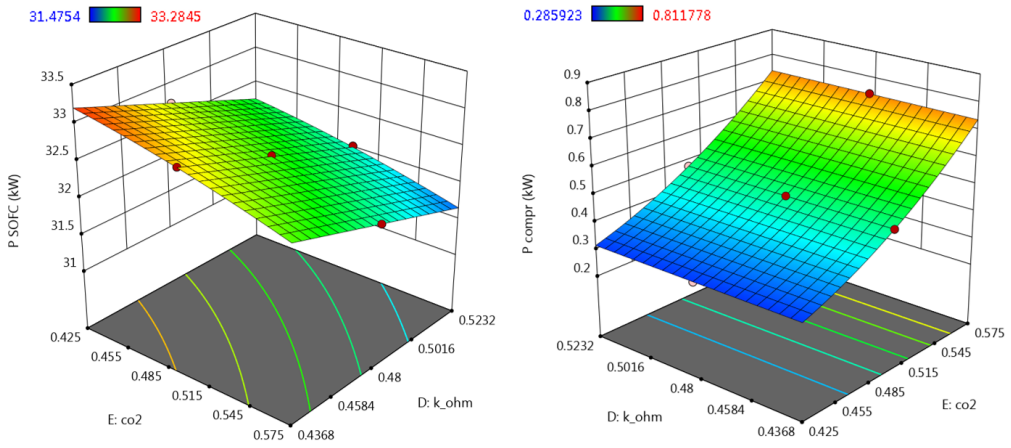


FIGURE 5.4: P_{SOFC} (left) and $P_{c,fuel}$ (right) response surfaces - UQ analysis on hybrid system

is always able to control the SOFC temperature, since its values are never equal to zero. Its opening is mostly influenced by η_t, and η_c, the higher these values are, the higher mass flow passes through the WG valve (figure 5.5). This effect is explained in more details in subsection 5.4.3.

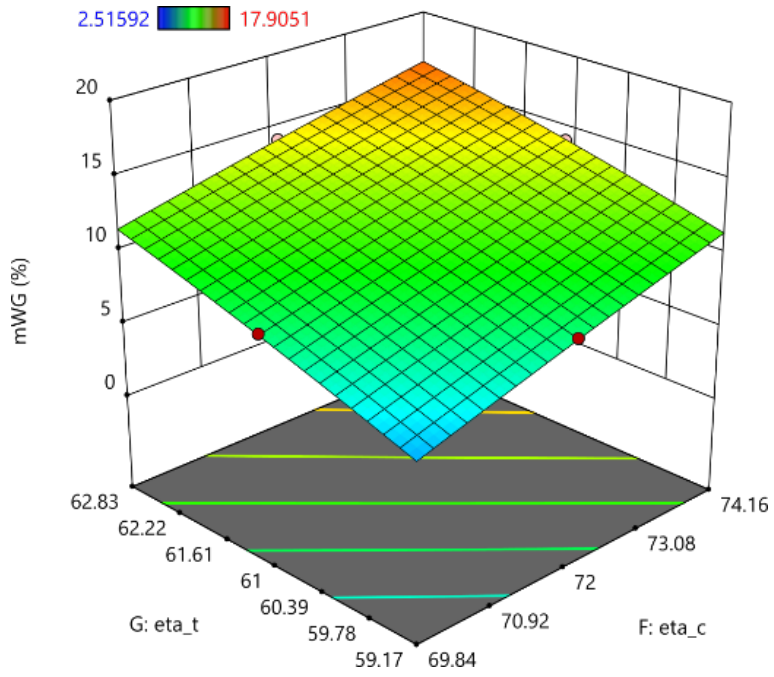


FIGURE 5.5: $\dot{m}_{WG}$ response surface - UQ analysis on hybrid system

STCR and shaft speed

The STCR varies between 2.20 and 2.47 (nominal is 2.38), remaining hence inside the safe region [150]. The STCR is mostly influenced by the CO2 percentage in the fuel and the ejector diameter with a direct proportionality (figure 5.6). In fact, the

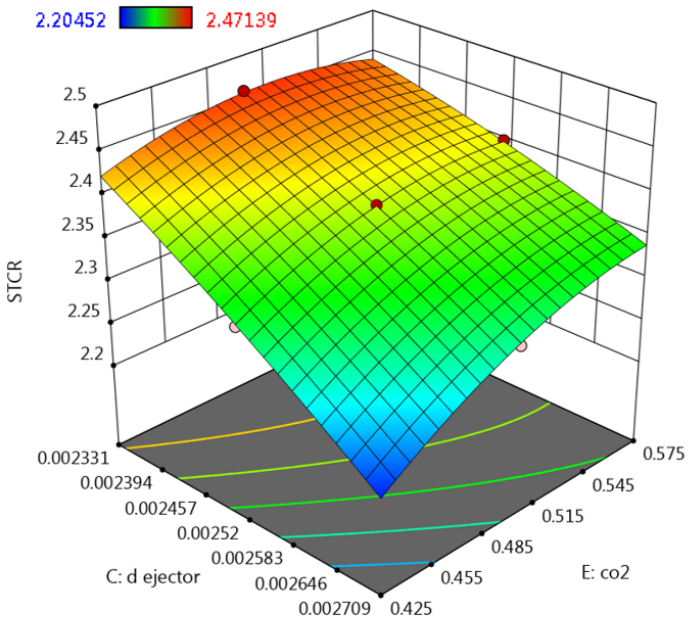


FIGURE 5.6: STCR response surface - UQ analysis on hybrid system

decreasing of the CO_2 percentage in the fuel composition results in an increase of the CH4 percentage, which, being at the denominator of the STCR equation (5.8), assuming other parameters of the equation not influenced, ends up with a decrease of the STCR. The pronounced curvature of the surface represents a significant importance of the quadratic terms. The STCR has a more linear dependence on CO_2 at lower dej values, while at high d_{ej} values the dependence of the CO_2 percentage increases significantly with a quadratic trend.
Regarding the ejector diameter, since the fuel mass flow at the primary anode ejector is maintained constant (as it depends only on fuel utilization factor U_f and electrical current, which are constant), a decrease in dej leads to a decrease of inlet-outlet pressure difference. Such difference leads to a decrease of the recirculation ratio and then to a decrease of the STCR. The turbocharger rotational speed can vary, considering the uncertainties analysed, between 258564 and 275714 rpm (nominal is 267853). Hence, the speed limit equal to 307 krpm, as provided by the supplier, is respected.

5.4.3 Analysis of the effects

The effects that each single input has on the response variables considered in this study are reported in figure 5.7, where the condition at the design point is taken as central point. The trend however persists even if another point of the design space is considered. For a qualitative analysis of the effects, table 5.6 presents with a simplified approach which is the effect of the increase of each input on the analysed output.

The pressure drops both at the hot and cold sides have a negligible influence on all the outputs analysed, even though such factors are significant for the response surface creation, as shown in table 5.6 . The effect of the input parameters on net power and net efficiency is the same, due to how the net efficiency is evaluated in

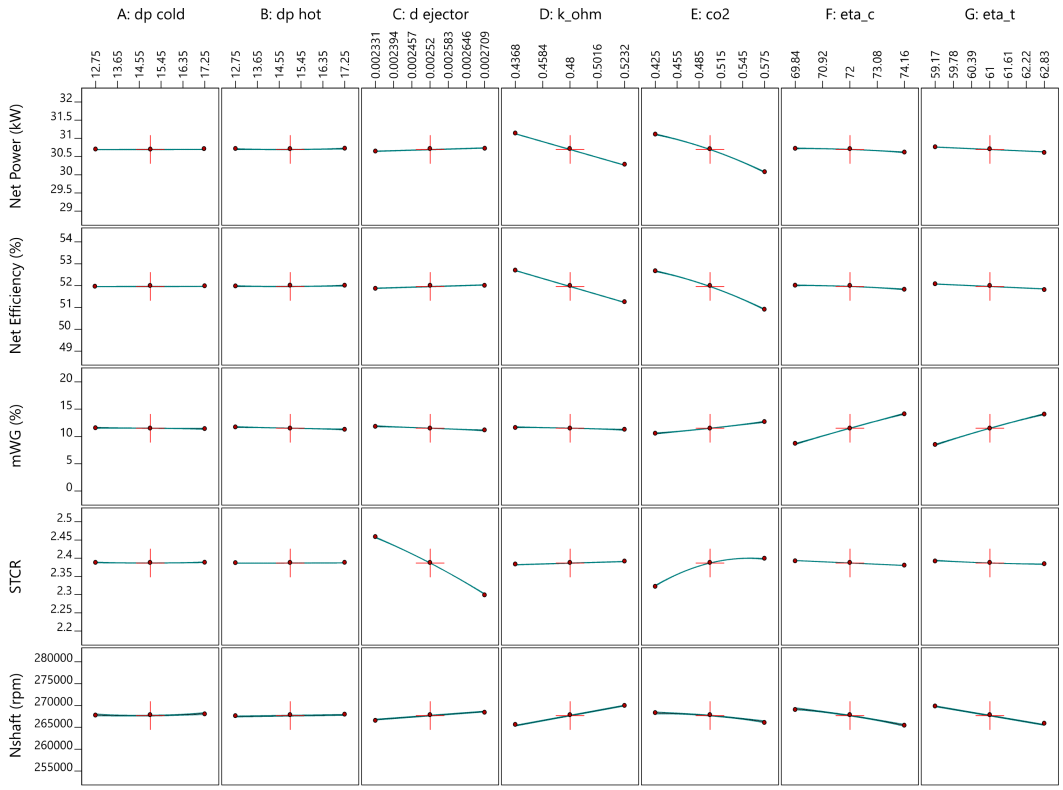


FIGURE 5.7: Analysis of the effects - UQ analysis on hybrid system

such systems (eq. 5.7) – i.e. function of net power and LHV of the fuel, which is a constant. These two outputs are mostly dependent on both ohmic losses and CO_2 percentage in the fuel. An increase of such parameters leads to a decrease of these two monitored outputs. A counterintuitive aspect to highlight is that, even if at small extent, an increase of the turbocharger efficiency (both on compressor and turbine sides) leads to a decrease of the power produced:

- If the compressor efficiency is higher, considering a certain pressure ratio, the outlet temperature is lower. Hence, it is necessary to open the WG valve to transfer more heat to the air flow and to comply with the fuel cell inlet temperature constraint, this reducing the system pressure.
- If the turbine efficiency is higher, more energy is extracted from the exhaust flow, whose temperature is lower. Since the air flow is heated by the turbine outlet flow, its temperature is also lower. To solve this issue, the WG valve must be opened as for the previous case, with same effect on fuel cell pressure.

In both cases, opening the WG valve reduces the turbine inlet mass flow and thus the mechanical energy transferred to the compressor. As a consequence, the compressor outlet pressure is lower and the power generated by the SOFC decreases.
The ohmic losses influence mostly the rotational speed. This is due because, if the ohmic losses increase, more heat is dissipated inside the cell. The recirculation ratio on the anode ejector decreases, hence more unburned fuel reaches the OGB. For this reason, more energy arrives at the turbine, which starts to increase the rotational speed. This effect, hence, increases the air mass flow that reaches the cathode side of

TABLE 5.6: Qualitative effect if each input is increased

	P_{net}	η_{net}	$\dot{m}_{WG}$	$STCR$	N_{shaft}
dp_{cold}	=	=	=	=	=
dp_{hot}	=	=	=	=	=
d_{ej}	↑	↑	=	↓	↑
k_{ohm}	↓	↓	=	↑	↑
CO_2	↓	↓	↑	↑	↓
η_c	↓	↓	↑	↓	↓
η_t	↓	↓	↑	↓	↓

the fuel cell, maintaining the cell set-point temperature.
The ejector diameter has a significant effect on the steam to carbon ratio. In fact, the model is developed in order to have always the same imposed fuel mass flow at the inlet; for these reasons the ejector diameter does not have effects on the power produced but only on the STCR, i.e. the ejector recirculation factor. The STCR increases by 4% if the ejector diameter is equal to the lowest value.

5.4.4 Interactions between variables

As already highlighted in the response surface analysis section, not only the single parameters have a significant effect on the response variable but also some interactions have a remarkable effect. The significant interactions are the ones that appear in the response surface equations. For what concerns the STCR, the interaction between the ejector diameter and the CO_2 percentage is the most important, even though the ohmic losses have an important effect as well. Looking at Figure 5.8 is it possible to note that at high value of ejector diameter the STCR is significantly influenced by the CO_2 percentage. On the other hand, at low level of ejector diameter, such influence is reduced significantly. In any case, the kohm influence remains almost the same: an increase of such parameter leads to an increase of STCR. The green points in next figures represent the point used for the response surface creation based on simulation model results.

Figure 5.9 depicts the interaction between k_{ohm} and d_{ej} on $\dot{m}_{WG}$ highlights that the $\dot{m}_{WG}$ is not particularly influenced by the ejector diameter, especial ly at high level of turbine efficiency. However, it can be observed that the ohmic losses have a significant effect on the mass flow that must pass through the WG valve at low compressor efficiency, even if such influence becomes negligible at high values of compressor efficiency.

Finally in Figure 5.10, the interaction between η_c and k_{ohm} on the net power is shown in Figure 9(a), from which it is possible to highlight that the influence of compressor efficiency on the net power is higher at high values of CO_2 percentage. In this case, the k_{ohm} has a linear influence on the response variable.

Thanks to the response surfaces, the extremes of P_{net} and η_{net} ranges can be found, which make it possible to have a broader view about the hybrid system behaviour. The highest P_{net} value is 31.66kW with a η_net of 54.0%. In this situation the WG valve is almost closed, since its mass flow is only 0.05% of the upstream value, with an STCR equal to 2.22 and with a turbocharger rotational speed N_{shaft} of 269458 rpm. The point where P_{net} and η_net are minimum can be also identified. In this condition

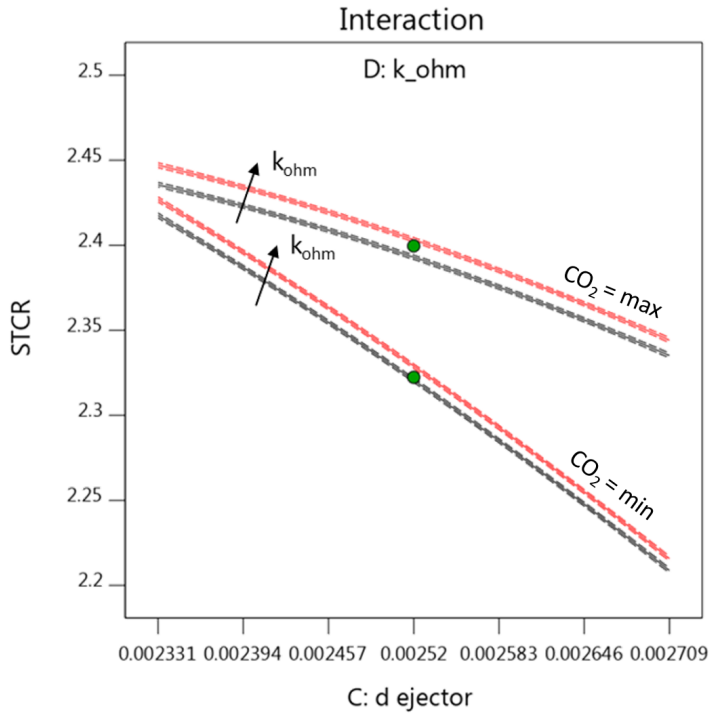


FIGURE 5.8: Interaction between k_{ohm} and d_{ej} on $STCR$ - UQ analysis on hybrid system

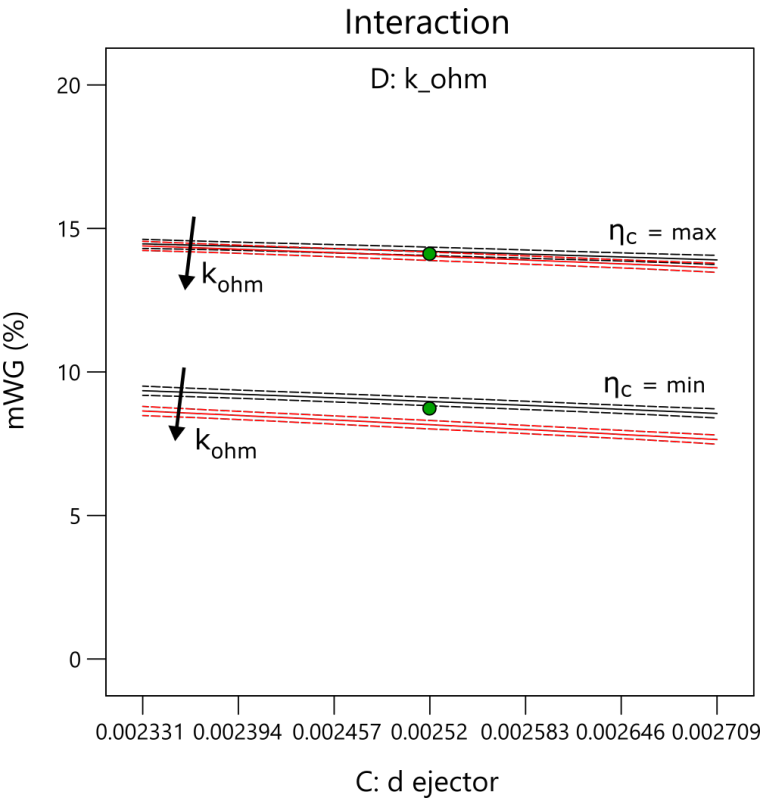


FIGURE 5.9: Interaction between k_{ohm} and d_{ej} on $\dot{m}_{WG}$ - UQ analysis on hybrid system

the system generates a P_{net} of 29.42 kW with a η_{net} of 49.8%; the WG valve bypasses

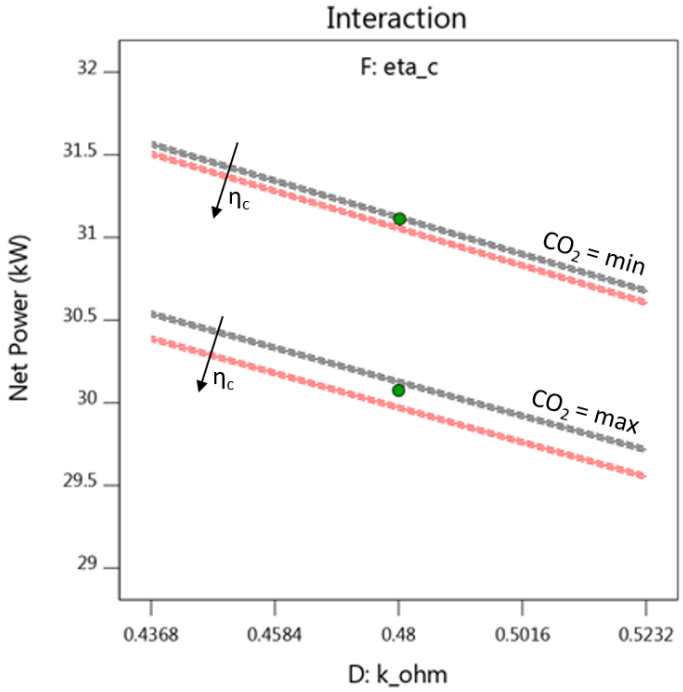


FIGURE 5.10: Interaction between η_c and k_{ohm} on P_{net} - UQ analysis on hybrid system

17.78% of the turbine upstream flow, the STCR is equal to 2.43 and Nshaft is 262746 rpm.

5.4.5 Monte Carlo simulation

The quality of the RS is verified prior of employing it for the Monte Carlo simulation (MCs) based probabilistic analysis [61]. MCs was applied to the obtained polynomial to evaluate the distribution of the outputs, considering 10^6 samplings. The number of samplings has been chosen to ensure the smallest mean square pure error (MSPE) possible [74]. The values of the input variables are extracted from normal distributions with the characteristics presented in Table 5.3. The creation of the RS is essential to carry out the MCS, because the time needed to run the simulation model 10^6 times wouldn't be acceptable. In fact, the time necessary to compute one of the RS results on the computer used during this work (CPU: Intel Xeon E-2176G 3.70Ghz, RAM: 16GB) is about 10^{-5} seconds, while a single run of the simulation model takes about 2 hours.
Figure 5.11 represents the obtained PDFs, while in Table 5.7 the mean, the standard deviation and the COV values are reported. The probability density functions of P_{net} (Figure 5.11a), N_{shaft} (Figure5.11b) and η_{net} (Figure 5.11c) resemble the shape of a Gaussian distribution while the STCR (Figure 5.11d) is more similar to a Weibull distribution.

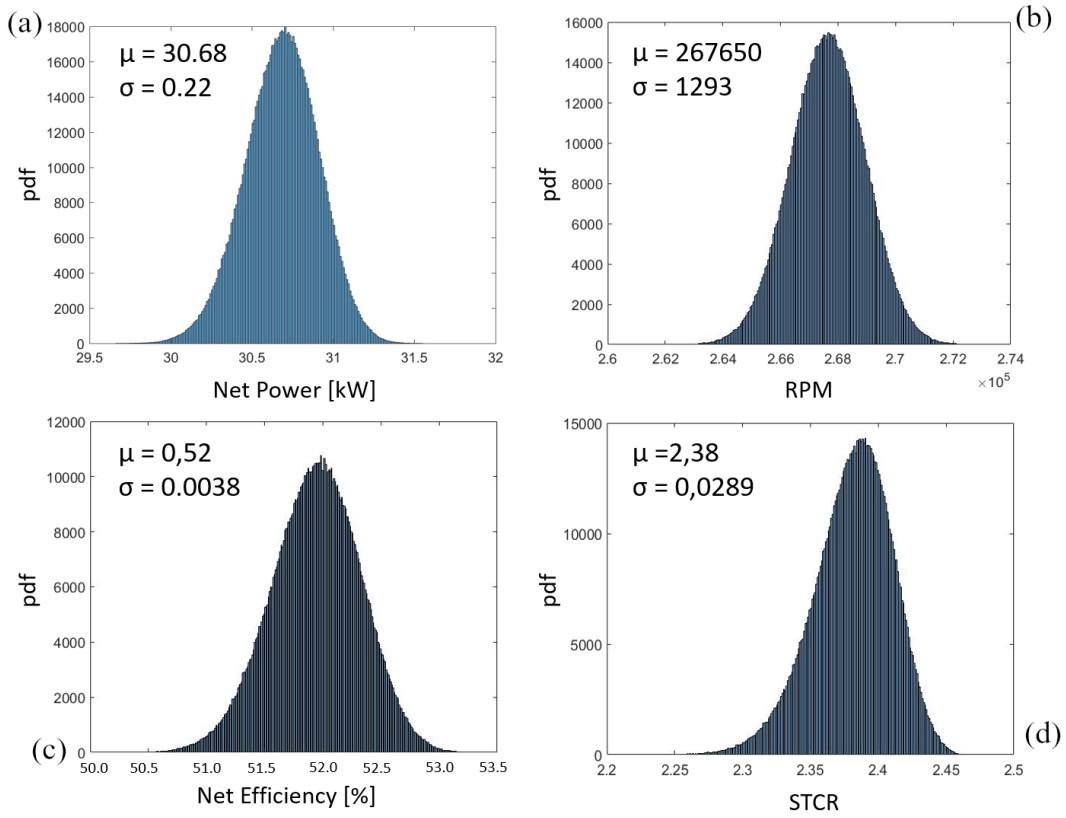


FIGURE 5.11: Monte Carlo simulation results - UQ analysis on hybrid system

5.5 Design under uncertainty

In this study [154], the effect of three parameters (factors) on the characteristics of the hybrid system in terms of performance and economic profitability within three different market scenarios is investigated. Such factors are (Table 5.14):

- Area of a single fuel cell (A_{cell})
- Exchange surface of the recuperator (A_{rec})
- Stack current density (J_{cell})

A 10% variation from the central point was chosen to determine the high and low levels of the factors. Adopting these levels, it is not necessary to change other components in the hybrid system to match the SOFC and the recuperator characteristics. A more substantial variation from the central point may require modifications of other components, significantly increasing the complexity of this analysis.

The uncertainties related to turbine and compressor maps in terms of efficiencies (η_t, η_c) and to the SOFC ohmic losses corrective coefficient (k_{ohm}) were also considered. A Gaussian probability density function was used to describe the uncertain parameters, as for engineering applications it has proven to be appropriate to describe stochastic variables when the probability density function (PDF) is not known

TABLE 5.7: statistical results of the output PDFs

Response	Unit of measure	μ	σ	COV
P_{net}	$[kW]$	30.68	0.224	0.73%
η_{net}	$[\%]$	0.52	0.0038	0.73%
$\dot{m}_{WG}$	$[\%]$	11.5	1.36	11.83%
STCR	$[-]$	2.38	0.0289	1.21%
N_{shaft}	$[rpm]$	267650	1293	0.48%

TABLE 5.8: Low and high levels considered for RS creation - Design under uncertainty

Factor	Name	Unit of measure	Central point	Low level	High level
A	A_{cell}	$[m^2]$	0.01278	0.01150	0.01406
B	A_{rec}	$[m^2]$	8.03	7.23	8.83
C	J_{cell}	$[A/cm^2]$	0.237	0.213	0.261

a-priori [40]. The coefficient of variance (COV) of stochastic parameters (Table 5.14), was determined based on industrial data and Authors' knowledge [152]. Compared to the work showed in section 4, it can be observed that in this case the COV related to the efficiencies is lower (5.3), that is because the turbocharger maps have been updated and improved and so the uncertainties have been reduced. Low values of turbomachinery efficiencies are justified by the low air flow processed (Table 5.1).

TABLE 5.9: Mean and coefficient of variance for the stochastic inputs - design under uncertainty

Variable	Unit of measure	μ	COV
η_c	$[\%]$	72	0.5%
η_t	$[\%]$	61	0.5%
k_{ohm}	$[-]$	0.48	3%

5.5.1 Response surfaces creation

As for the study on the propagation of uncertainties (section 4), it was used the Central Composite Design Face Centered (CCF) method to build the response surfaces of the model. To consider the uncertainties within the CCF, each run was performed 15 times (n_{reps}), taking for each simulation a random value of the uncertain parameters, according to the PDF. Thus, the total number of design points has become:

$$n_{runs} = n_{reps}\left(2^k + 2k + 1\right) \tag{5.11}$$

which for k=3 and n_{reps}=15 ends up in a total of 225 runs. The response surfaces were created for the net power (P_{net}), the net efficiency (η_{net}), the Internal Rate of

Return (IRR), PayBack Period (PBP) and the capital cost of the plant (C_{plant}). The total initial investment was computed as a function of the costs of the main components of the hybrid system [155, 123, 156]. In particular, the recuperator cost was computed according to [123] as:

$$C_{REC} = 1.5 \cdot \left(r_1 \dot{m}_{in}^{cold} \left(p_{in}^{cold} \right)^{-0.5} \left(\Delta p^{-0.5} \right) \frac{\epsilon_{REC}}{1 - \epsilon_{REC}} \right) f \tag{5.12}$$

where:

- r_1 is a constant
- $\dot{m}_{in}^{cold}$ is the cold side inlet mass flow
- p_{in}^{cold} is a the cold side inlet pressure
- Δp is the percentage pressure drop
- ϵ_{REC} is the recuperator effectiveness
- f is a parameter dependendant on the recuperator material.

while the SOFC stack cost was computed according to [155] as:

$$C_{SOFC,stack} = C_{SOFC} + C_{inv} + C_{aux} \tag{5.13}$$

where:

$$C_{SOFC} = A_{cells} \left(2.96 T_{cell} - 1907 \right) \tag{5.14}$$

$$C_{inv} = 10^{-5} \left(\frac{P_{SOFC}}{500} \right)^{0.7} \tag{5.15}$$

$$C_{aux} = 10\% C_{SOFC} \tag{5.16}$$

of which A_{cells} is the total area of the fuel cells, T_{cell} is the SOFC average temperature and P_{SOFC} is the SOFC power design value.
The reformer cost was instead calculated as [155]:

$$C_{REF} = 130 \left(\frac{A_{REF,fin}}{0.093} \right)^{0.78} + 3240 V_{REF}^{0.4} + 21280.5 V_{REF} \tag{5.17}$$

where $A_{REF,fin}$ is the reformer finned aread and V_{REF} is the reformer volume. The cost of piping and other components is estimated equal to 10% of the total plant cost. The price of the turbocharger has been assumed equal to 800€after confrontation with professionals in industry.
All the costs have been updated to the year of reference of this work (2018) using the Chemical Plant Cost Indexes (CEPCI) [157].
The cash flow of the j-th year (CFN_j) was computed as:

$$CFN_j = C_{el} P_{net} - C_{fuel} P_f - C_{maint} \tag{5.18}$$

where the annual maintenance was computed as the sum of a 10% of the SOFC initial investment and a 3% of the initial investment of the other components, considering a replacement of the stack after 5 years, amortized along the 10 years life span of

the plant. The value of C_{fuel} is computed from the cost of the natural gas CH_4, taking into account its content in the biogas (50% molar composition) and assuming no extra cost due to the CO_2 in it. The economic profitability of the system was computed considering three different market scenarios: Italy, United Kingdom and United States (New York). For each of these countries a different cost of electricity C_{el} and fuel C_{fuel} was evaluated, according to the size of the plant, its consumption and taking 2018 as reference year. The system was assumed to operate for 8000 Effective Operative Hours (EOH) per year. The main economic parameters used for the analysis are reported in the following table (Table 5.10).
Firstly, an analysis of variance (ANOVA) of the responses was performed to identify

TABLE 5.10: Main economic parameters - Design under uncertainty

Revenue	Italy (IT)	$C_{el} = 0.1636\,€/kWh$	[158]
	United Kingdom (UK)	$C_{el} = 0.1598\,€/kWh$	[158]
	New York (NY)	$C_{el} = 0.1364\,€/kWh$	[159]
Operational Cost	Italy (IT)	$C_{CH_4} = 0.0437\,€/kWh$	[160]
	United Kingdom (UK)	$C_{CH_4} = 0.0279\,€/kWh$	[160]
	New York (NY)	$C_{CH_4} = 0.0215\,€/kWh$	[161]
Maintenance cost	$C_{maint} = 10\%C_{SOFC} + 3\%\left(C_{plant} - C_{SOFC}\right)$		
Operating hours	$EOH = 8000h/yr$ $EOL = 10yrs$		

the most suitable model with Design-Expert® 12.0. The model was selected with the p-value method, including the terms with a p-value < 0.001, using a forward selection. The forward selection of Design-Expert works by putting all the terms in the model one at a time, adding to the model only the terms which improve the criterion. Terms not yet in the model are checked until no further improvement can be found given the terms already in the model. This selection has led to the choice of two main models: reduced two-factor interaction (2FI) and reduced quadratic, as reported in Table 5.11. The generalized 2FI and quadratic equations can be expressed as:

$$y = \sum_{i,j=1}^{k} a_i x_i + b_i x_i x_j \quad \begin{cases} \text{2FI} & \forall i,j \in \mathbb{N} s.t. i \neq j \\ \text{quadratic} & \forall i,j \in \mathbb{N} \end{cases} \tag{5.19}$$

while "reduced" means that not all the terms of the 2FI and quadratic model were considered. For the analyzed responses the reduced 2FI preserved all the first order terms and, amongst second order terms, only the product between term B and C (J_{cell} and A_{REC}). Also the quadratic models included all the first order terms, but considered B2 for the PBP and BC and C^2 when evaluating the net efficiency. In Table 5.11 also the values of the predicted and adjusted R^2 are reported. As rule of thumb, predicted and adjusted R^2 should be within 0.2 of each other, otherwise there could be a problem with either the data or the model [162].

From Table 5.11 it can be observed that the adjusted and predicted R^2 of the responses are within 0.2 each other and also that they all present high values of R^2, well above 0.9, indicating that the models fit simulation data extremely well and can be used to properly navigate the design space. Only the PBP and IRR of Italian case

TABLE 5.11: Selected models for RS and R^2 of the responses - Design under uncertainty

Response	Model	R^2_{adj}	R^2_{pred}
P_{net}	Reduced 2FI	0.9963	0.9962
η_{net}	Reduced quadratic	0.9478	0.9463
C_{plant}	Reduced 2FI	0.9988	0.9988
PBP_{IT}	Reduced quadratic	0.9065	0.9046
PBP_{UK}	Reduced quadratic	0.9759	0.9753
PBP_{NY}	Reduced quadratic	0.9817	0.9809
IRR_{IT}	Reduced 2FI	0.9084	0.9064
IRR_{UK}	Reduced 2FI	0.9741	0.9735
IRR_{NY}	Reduced 2FI	0.9797	0.9792

show a R^2 close to 0.9, indicating that some higher (but acceptable) inaccuracy exist in such a case, and should be considered in the analysis of results.

5.5.2 Analysis of the responses

Having established the quality of the response surfaces, the same were used in the first place to evaluate within which range the outputs can vary. From Table 6 it can be observed that the plant can have a wide range of performance within the design space considered. The net power can vary of about 9kW, and the efficiency of about 4 percentage points between the minimum and the maximum values obtainable. On the economic side, the capital cost of the plant is also subject to a consistent variability, while analyzing the PBP and the IRR it can already be noticed that the plant would be more profitable in United Kingdom amongst the countries considered, with a PBP which ranges between 2.2 and 2.7 years. This is directly related to the quite high cost of electricity (Table 5.10), self-consumed by the plant owner (eq. 5.18), and to the low cost of fuel. The extremely low cost of natural gas of New York contributes to make the US solution interesting from an economic point of view, however the low cost of electricity impacts directly the annual cash flow, making that solution good but less profitable than the UK one. Finally, the high cost of natural gas in Italy (Table 5.12) impacts directly the PBP and the IRR, making this solution interesting as the PBP is of about 3.5 years, but less profitable than the other two scenarios considered.

In the following response surface figures, the colored dots represent the simulation points used for the RS creation, based on model results (darker dots are above the surface, lighter below).

Performance: net power and net efficiency

As expected, the ANOVA suggested that the factors which have the strongest impact on the net power are the current density and the cell area, as they directly impact the size of the stack and then the power production (Figure 5.12). In fact, the highest P_{net}

TABLE 5.12: Selected models for RS and R^2 of the responses - Design under uncertainty

Response variable	Unit of Measure	Nominal value	Minimum value	Maximum value
P_{net}	kW	30.71	25.36	36.08
η_{net}	–	0.5198	0.4925	0.5396
C_{plant}	k€	50.655	46.396	54.962
PBP_{IT}	years	3.34	3.06	3.66
PBP_{UK}	years	2.33	2.23	2.73
PBP_{NY}	years	2.66	2.41	2.98
IRR_{IT}	–	0.2724	0.2417	0.3040
IRR_{UK}	–	0.4150	0.3471	0.4371
IRR_{NY}	–	0.3575	0.3136	0.4014

values are obtained for the maximum size of the SOFC stack.
When considering the net efficiency, the ANOVA suggests also a strong dependence

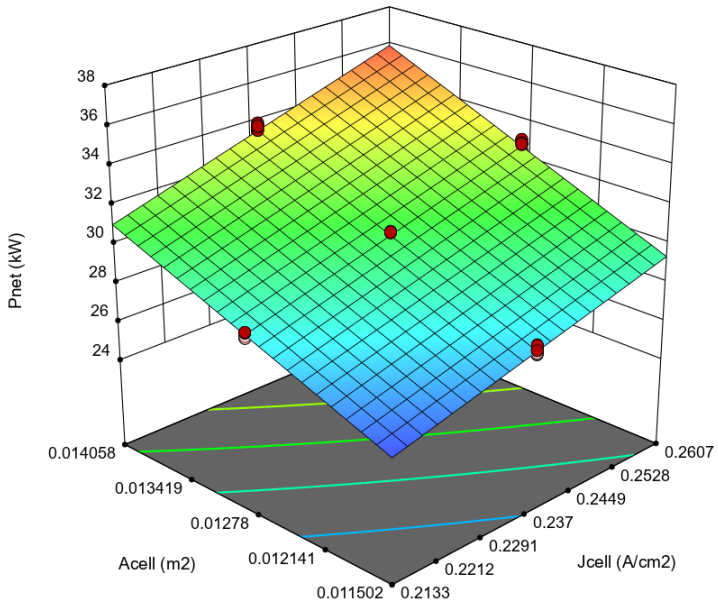


FIGURE 5.12: Net power response surface - design under uncertainty

on current density and cell area, particularly the current density is taken into higher consideration, as it is also present as second order term for the quadratic model. From Figure 5.13 it can be observed that the net efficiency presents a maximum for low current density, while when considering a fixed J_{cell} the cell area has a very low impact on the resulting efficiency. This is due to the ohmic losses, as they are directly related to the current: a lower stack current would guarantee lower ohmic losses and then a higher efficiency. It is interesting to highlight that for this RS the colored dots corresponding to the model simulation are more spread, resulting then in a lower R^2 as reported in Table 5.11.

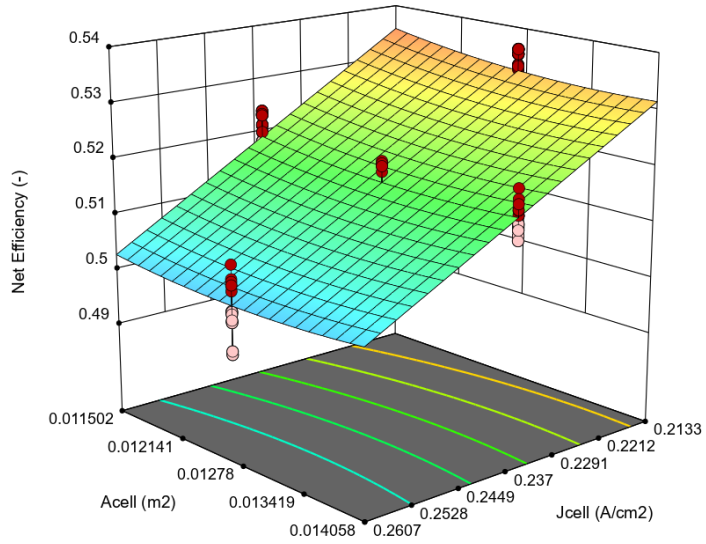


FIGURE 5.13: Net efficiency response surface - design under uncertainty

Economic parameters: capital cost, PBP and IRR

The capital cost of the plant is dependent mostly on the size of the SOFC stack (Figure 5.14), as this component is more expensive than the recuperator.

The payback period presents a similar behaviour for all the scenarios considered

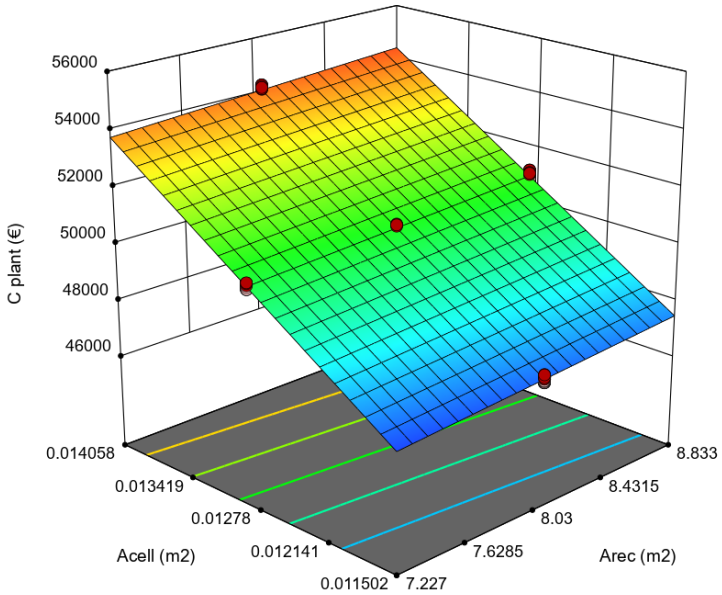
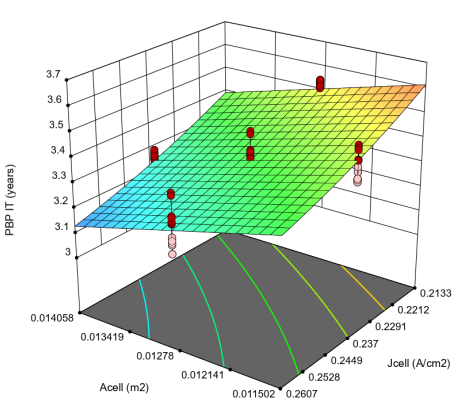


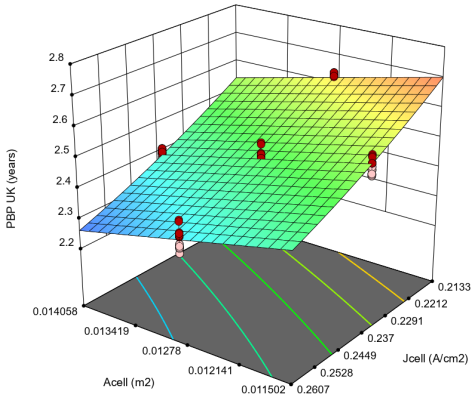
FIGURE 5.14: Capital cost response surface - design under uncertainty

and is mostly influenced by the current density and the cell area. The current density factor is also considered as second-order term (B^2) within the reduced quadratic model chosen (Table 5.11). From Figure 5.15 it can be observed that the behaviour of the RS is the same for all the countries considered. It can then be highlighted that

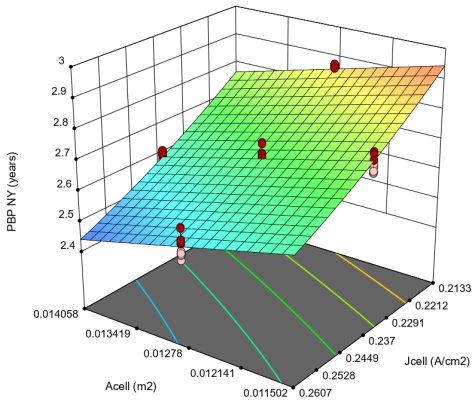
the lowest PBP is obtained when considering a bigger SOFC stack, corresponding to high values of cell area and current density. These parameters have in fact a similar influence on the net power (Figure 2), impacting directly the revenues of the plant and so the PBP.
The internal rate of return is also mostly influenced by the SOFC stack size, from



(A) Italy



(B) United Kingdom



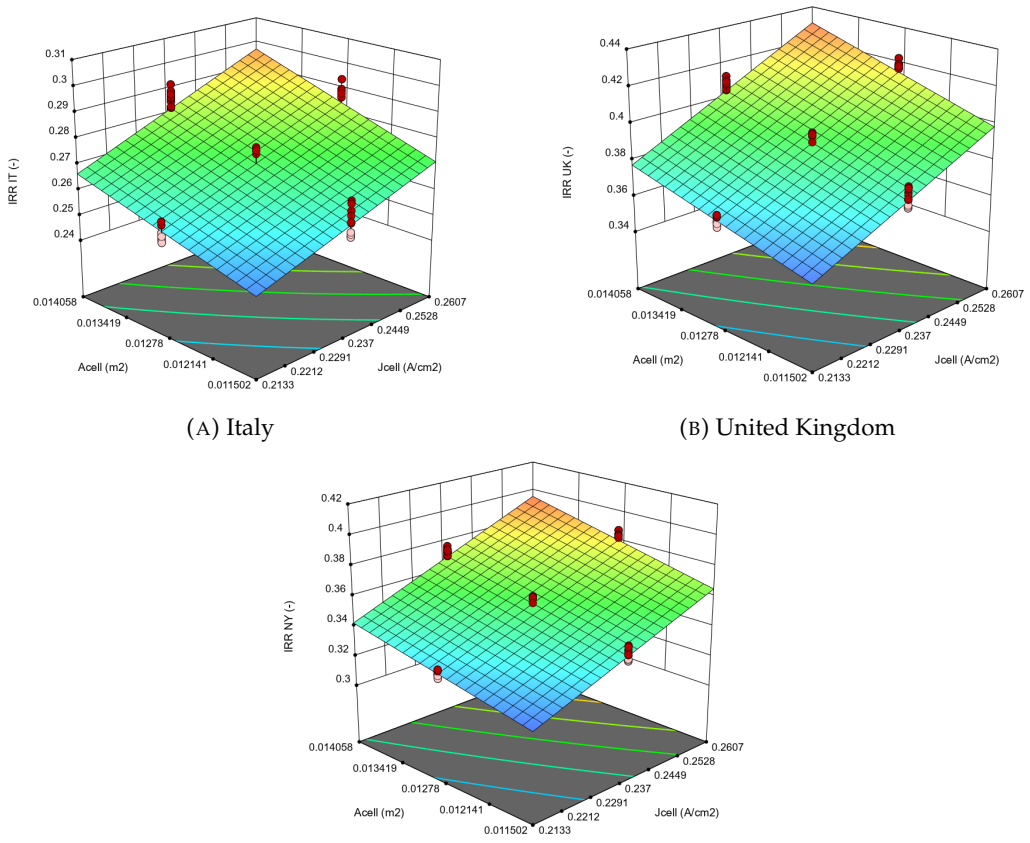
(C) New York

FIGURE 5.15: Payback period response surfaces - design under uncertainty

the single factor B and C and from their product (Figure 5.16). Similar conclusion to the PBP can be drawn for the IRR as it concerns the dependence on the two factors.

5.5.3 Multi-objective optimization

The polynomial responses of the net power and of the plant cost were used for a multi-objective optimization with genetic algorithm [163], using the three factors enlisted in Table 5.14 as decision variables. The high level and low level considered for the RS creation were used as upper and lower bounds respectively, as the polynomial RS is only valid within those boundaries.
The number of generations for the genetic algorithm was set in order to obtain a small average spread value and a minimum variation of it amongst the generations. The spread is a measure of the movement of the Pareto set: a small value means that the extreme objective function values do not change much between iterations, and



(A) Italy

(B) United Kingdom

(C) New York

FIGURE 5.16: Internal rate of return response surfaces - design under uncertainty

that the points on the Pareto front are spread evenly.
In multi-objective optimization, the Pareto front is a set of optimal, non-dominated solutions. A solution is considered non-dominated if no improvement can be achieved in one of the objectives without leading to a degradation of at least one of the other objectives [164]. All the solution which do not lie on the Pareto front are considered instead dominated, or locally optimal.
The resulting Pareto front is reported in Figure 5.17. The first thing that can be noticed by plotting the design point on the figure, is that the Design Point is a dominated solution. As a matter of fact, the realization of a hybrid system able to guarantee the same net power of about 30kW could be done with a capital cost investment of about 49k€instead of 50.6k€. This could be obtained by using a similar recuperator (8.05 m^2 instead of 8.03 m^2), and a SOFC stack with a slightly higher current density (0.257 A/cm^2 instead of 0.237 A/cm^2) and a similar but smaller fuel cell area (0.01223 m^2 instead of 0.01278 m^2). On the other side, with the same investment in terms of capital costs a 32kW plant could be built by optimizing the SOFC stack size and the recuperator area.
Figure 5.17 also highlights that the two points of the Pareto front at maximum P_{net} and maximum capital cost are quite close one to each other. It was then decided to investigate further these two points: if the investor had a higher amount of money to invest it should be evaluated which one of these two solutions would be more robust, guaranteeing a more stable net power production when considering the uncertainties in turbine and compressor maps and k_{ohm}. The characteristics of the two

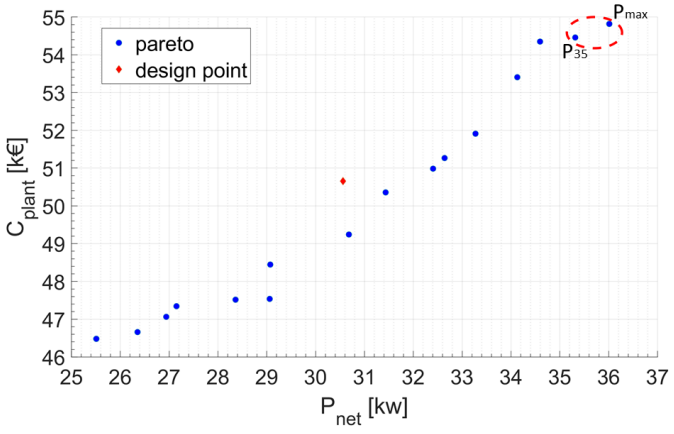


FIGURE 5.17: Pareto front (red dot: design point) - design under uncertainty

highlighted points are reported in Table 5.13.

TABLE 5.13: Maximum power Pareto front solutions - design under uncertainty

Case	P_{net} [kW]	C_{plant} [k€]	A_{REC} $[m^2]$	J_{cell} $[A/cm^2]$	A_{cell} $[m^2]$
P_{max}	36.08	54.812	8.833	0.2607	0.014058
P_{35}	35.61	54.395	8.468	0.2596	0.013964

5.5.4 Robust design

To perform a proper analysis of the points highlighted in Figure 5.17, two new response surface have been created using CCF method, considering the values of A_{REC}, J_{cell} and A_{cell} reported in Table 5.13 and varying the uncertain parameters defined before (Table 5.14). The RS were created considering that low and high levels had a range equal to $\pm 3\sigma$. Such range was considered since in a normal distribution this interval contains 99.73% of the data (Table 8).

The ANOVA was performed to obtain the most suitable model to represent net

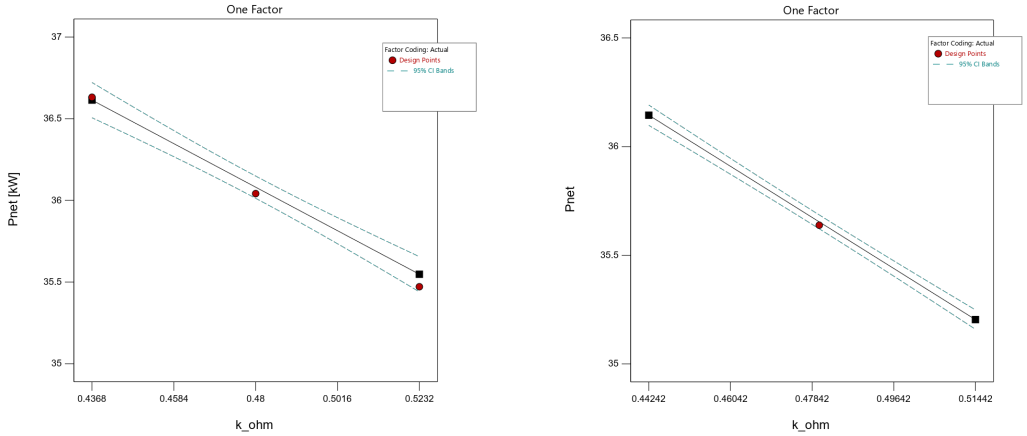
TABLE 5.14: Factor levels used for the robust design RS - design under uncertainty

Factor	**Name**	**Unit of measure**	**Central point**	**Low level**	**High level**
D	η_c	[%]	72	70.5	73.5
E	η_t	[%]	61	59.5	62.5
F	k_{ohm}	[−]	0.48	0.4368	0.5232

power and net efficiency, selecting terms with p-value < 0.001 and forward selection

as in section 5.1.
A simple linear model was chosen to represent both efficiency and power, with an R^2 above 0.95 for all the cases. As expected, the k_{ohm} factor had the strongest impact on the responses for both cases analysed (Table 5.13), as the power and efficiency are directly influenced by the SOFC performance, while the turbocharger performance impact those terms only indirectly. The one-factor graph for is represented in Figure 5.18, where the response of the net power is plotted against k_{ohm}. From this figure it can be noticed that the red dots, representing the points simulated with the Matlab-Simulink model, lie in the 95% confidence band. As expected, high ohmic losses results in a lower SOFC power production which are directly reflected on the overall net power.

The equations representing the models obtained from the ANOVA analysis were

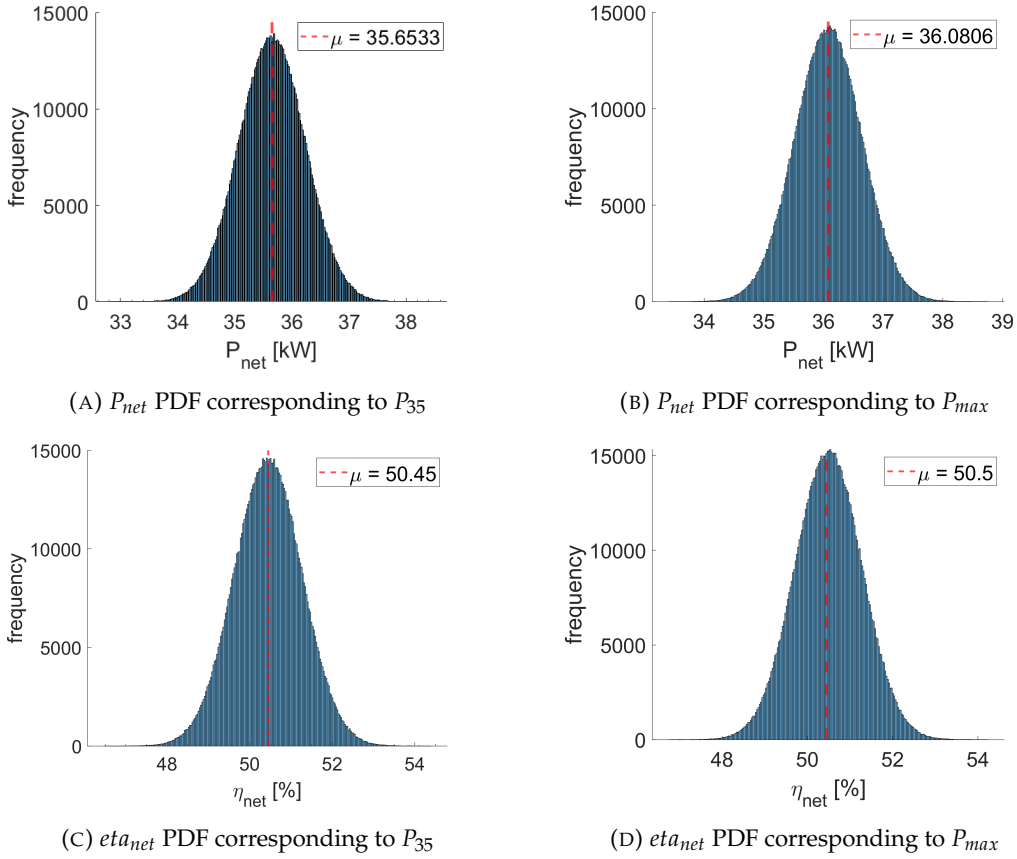


(A) P_{max} (B) P_{35}

FIGURE 5.18: One factor model representation of P_{net} vs. k_{ohm} - design under uncertainty

then used to perform a Monte Carlo simulation (MCs) to evaluate the PDF of the responses and their standard deviation. MCs was applied considering 10^6 samplings. The number of sampling was chosen to ensure the smallest mean square pure error (MSPE) possible [74]. The values of the input variables, which are the factors used to build the RS (Table 5.14), were extracted from the corresponding normal distributions.
In Figure 5.19, the PDF of the net power and the net efficiency of the two cases reported in Table 5.13 can be observed. As expected from the distribution of the inputs (Gaussian), also the outputs present a Gaussian distribution. However, it is worth to highlight that the mean value (μ) of the P_{max} case net power (Figure 5.19B) is exactly the one computed from the multi-objective optimization performed in section 5.3 (Table 5.13), while the mean value of net power for the P_{35} case (Figure 5.19A) is slightly different (0.11% higher). However, this difference can be considered negligible, meaning that the PDFs obtained from the MCs are representative of the output behaviour.

Analyzing then the coefficient of variance of the outputs (COV), it can be noticed that the P_{35} case presents slightly higher COV both for net power and net efficiency (Table 5.15). This means that the variance of this case is higher and so it is less robust. Because of this, the system may be operating in a wider range of performance, which is usually not a beneficial condition. From these results and from the ones in Table 5.13 it can then be concluded that, with a capital cost investment slightly



(A) P_{net} PDF corresponding to P_{35}

(B) P_{net} PDF corresponding to P_{max}

(C) eta_{net} PDF corresponding to P_{35}

(D) eta_{net} PDF corresponding to P_{max}

FIGURE 5.19: Monte Carlo simulation outputs for the robust design - design under uncertainty

higher compared to the P_{35} solution (i.e. 54.4 k€vs. 54.8 k€), P_{max} design guarantees a net power of 36 kW with a power production less sensible to uncertainties in the turbocharger maps and in the SOFC ohmic losses.

The economic parameters, which characterize the chosen solution at design point,

TABLE 5.15: COV of the monitored outputs - design under uncertainty

Case	P_{net}			η_{net}		
	μ	σ	COV	μ	σ	COV
P_{max}	36.08	0.56	1.55 %	50.50	0.79	1.56%
P_{35}	35.65	0.58	1.89%	50.45	0.82	1.62%

are listed in Table 5.16. As highlighted in section 5.1, the best-case scenario is the one in UK, featuring the lowest payback period and an exceptionally high IRR of about 0.45, due to the low costs of the fuel and to the mid-high costs of electricity. The NY case appears also interesting, with a PBP lower than 2 years and a half, thanks to the low cost of the fuel. On the other hand, the Italian scenario appears to be the worst for such kind of plant due to the high costs of the natural gas. However, the PBP of about 3 years is still interesting from an economic point of view.

TABLE 5.16: Selected models for RS and R^2 of the responses - Design under uncertainty

Output	Unit of	Nominal value
C_{plant}	k€	54.812
PBP_{IT}	years	3.13
PBP_{UK}	years	2.14
PBP_{NY}	years	2.44
IRR_{IT}	–	0.2977
IRR_{UK}	–	0.4552
IRR_{NY}	–	0.3952

5.6 Off-design analysis under uncertainty

To perform a study of the off-design of the hybrid system described in section 5.3, the results of the responses obtained section 5.5 have been used, in particular focusing on the Italian market scenario (table 5.10 and taking as starting design the one which guaranteed the highest IRR (figure 5.16A). This layout would be obtained using a SOFC stack with an A_{cell} of 0.01406 m^2, a current density of 0.261 A/cm^2, and a recuperator with an exchange surface of 8.83 m^2, which allows to achieve the maximum net efficiency (54% and power output (36kWe) [42].
Starting from this design, an off-design analysis of the system was performed by analysing the behaviour of the plant at 100%, 85% and 70% of the nominal load. The electric power generated by the SOFC is directly related to the current density, therefore it is possible to follow the power load demand varying the current density value proportionally to it. The fuel is proportionally varied, keeping constant the utilization factor at the design value of 0.8.

5.6.1 Analysis of the responses

Two new RS have been created, for each load, using the uncertainties enlisted in Table 5.14 as factors and considering a low and high level equal to $\mu \pm 3\sigma$ for each factor. The reason behind this choice is that, as stated before, the three uncertain parameters were considered distributed according to a Gaussian, so within the range of $\pm 3\sigma$ lie 99.7% of the values and it is then possible to capture properly the behaviour of the system.
The response surfaces of net power and net efficiency were created, using the CCF method with a number of repetitions, which in this case was equal to 1 (eq. 5.11), as the uncertainties became the actual factors of the RS.
ANOVA suggested that for net power and net efficiency a linear model was adequate to create a response of both net power and net efficiency. From Table 5.17 it can be observed that the R^2 is almost equal to 1 for net power and net efficiency at 85% and 70% load. This is however different when considering 100% load, where both predicted and adjusted R^2 have a lower, but acceptable, value. This is because the simulation point with minimum η_c and η_t and maximum k_{ohm} deviates from the standard behaviour of the system, reaching a SOFC average temperature of about

1141K, which is 8K higher than the maximum SOFC temperature (1133K) , with the WGV closed and so without any possibility of cooling the stack. For this reason, the extreme points of the 100% load responses should be considered with caution.

The results showed in the last three columns of Table 5.17 have been reported in

TABLE 5.17: Off-design analysis results

	Response	R^2_{pred}	R^2_{adj}	Nom. value	Min. value	Max. value
100%	P_{net} [kW]	0.8106	0.8994	35.97	35.23	36.65
	η_{net} [−]	0.8106	0.8994	0.5034	0.4931	0.5129
85%	P_{net} [kW]	0.9999	0.9999	32.07	31.62	32.54
	η_{net} [−]	0.9999	0.9999	0.5281	0.5206	0.5357
70%	P_{net} [kW]	0.9999	0.9999	27.72	27.44	27.99
	η_{net} [−]	0.9999	0.9999	0.5543	0.5487	0.5597

Figure 5.20. From this figure it can be highlighted already that the variability of net power and net efficiency is slightly higher at 100% load and decreases while lowering the load.

From figure 5.21 and 5.22 it can be observed that the behaviour of the response

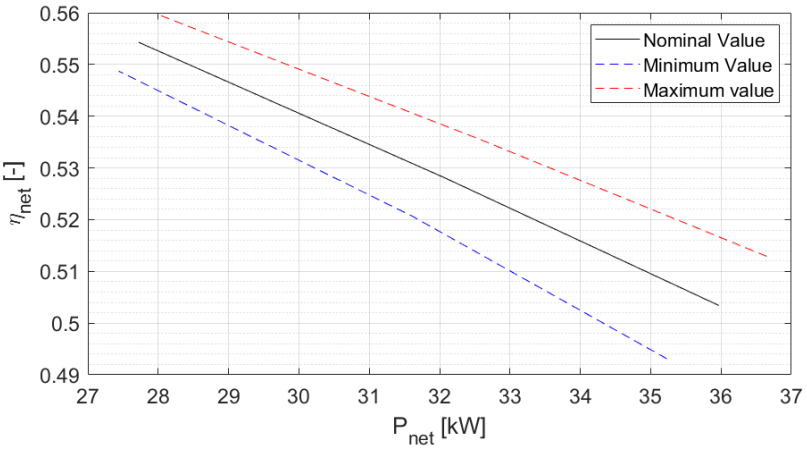
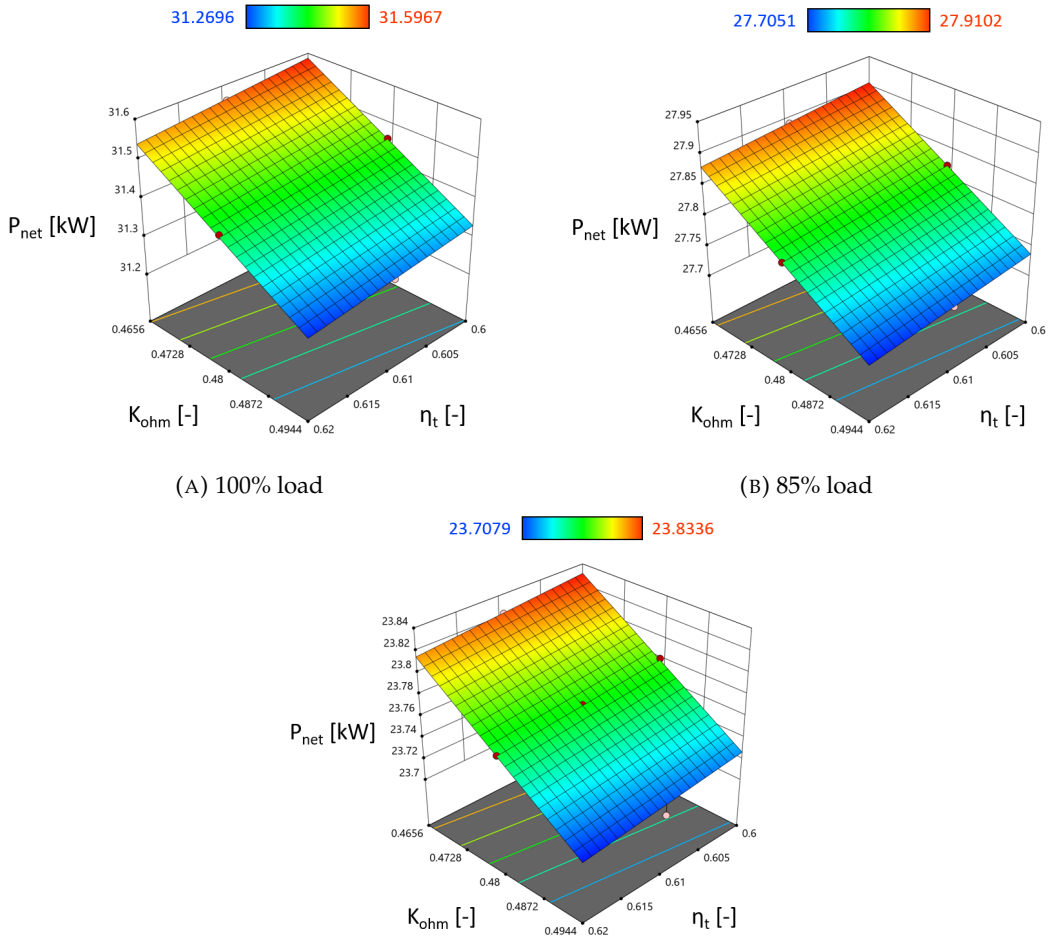


FIGURE 5.20: Net Power and net efficiency variability at different loads (T_{amb} = 15C) - off-design analysis

surfaces is similar between the considered loads, as expected. Moreover, it can be pointed out that both net power and net efficiency are mostly dependent on the ohmic losses, as they impact directly the power generated by the fuel cell stack. In particular, for small values of k_{ohm} the ohmic losses are lower and so the P_{SOFC} is higher (eqs. 5.6,5.5), leading to higher values of net power and net efficiency. The efficiency of the turbine and compressor have a smaller impact on the considered responses, as the turbocharger in this plant drives the system but does not generate power directly.



(A) 100% load

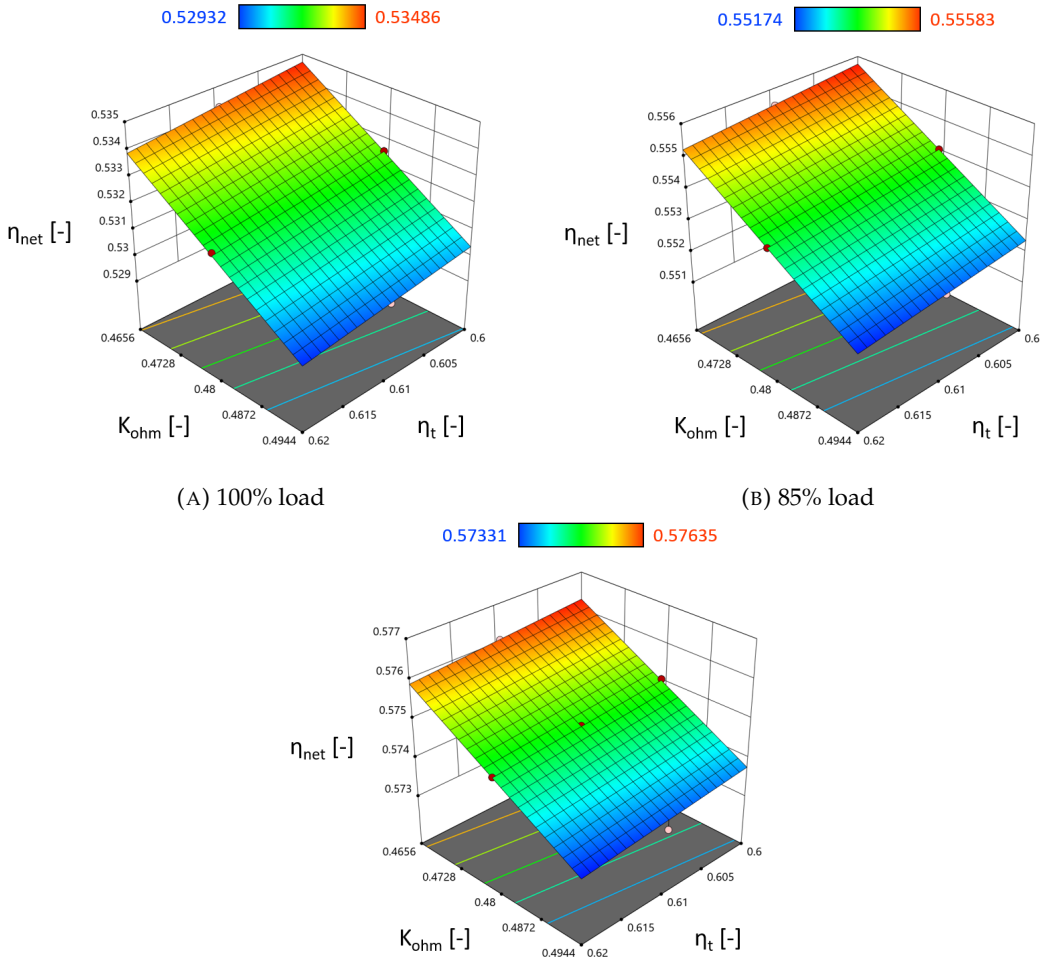
(B) 85% load

(C) 70% load

FIGURE 5.21: P_{net} response surfaces at different loads - off-design analysis

5.6.2 Monte Carlo simulation

The polynomial responses were then used to perform a Montecarlo simulation on the system, assuming that the system in real conditions would work for one third of the operating hours at each different load. The Italian economic assumptions done in section 5.5 and enlisted in Table 5.10 were used for this simulation, considering the uncertain parameters distributed according to a Gaussian PDF with the mean and standard deviation reported in Table 5.14.
For the Montecarlo simulation also the uncertainties related to the economic parameters, in terms of electricity and fuel cost, were considered. In particular, the values reported in Table 5.5 were used to represent the mean of the Gaussian PDF, while a coefficient of variance of 3% was taken into account, leading to a maximum variation of the prices of $\pm 9\%$ from the mean ($\pm 3\sigma$). It was assumed that the prices of fuel and electricity were uncorrelated, which could be possible in a scenario with increasing renewable energy sources.
The Montecarlo simulation was performed with $2.5 \cdot 10^5$ samples, in order to guarantee the lowest mean square pure error possible, analysing the impact of the uncertainties on the revenues of the plant in terms of payback period (PBP) and internal



(A) 100% load

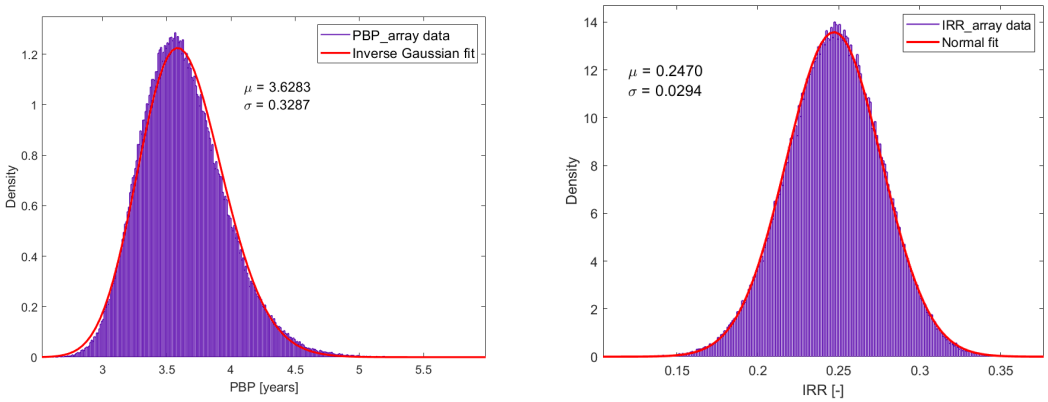
(B) 85% load

(C) 70% load

FIGURE 5.22: η_{net} response surfaces at different loads - off-design analysis

rate of return (IRR) within the Italian market scenario. As said before, the creation of the RS is essential to carry out the Montecarlo simulation, as the time needed to run the simulation model $2.5 \cdot 10^5$ times would not be acceptable.
Analyzing the resulting PDFs reported in Figure 5.23, it can be observed that they present two different shapes. As a matter of fact, the PBP distribution (Figure 5.23A) is quite well represented by an Inverse Gaussian, while the IRR (Figure 5.23B) by a Normal distribution. It is worth mentioning that the tail values of both parameters are due to the uncorrelation of the prices. In particular, the lowest IRR are the results of a combination of low fuel prices and high electricity prices, while the opposite combination leads to high IRR. The mean of the PBP is acceptable, considering the 10 years life of the plant, as well as the coefficient of variance (COV = 9%), however to further reduce the variability a careful evaluation of the evolution of prices along the years should be considered when performing a proper feasibility analysis of such plant.
Moreover, if measures should be undertaken at design level to accommodate for potential uncertainties in the actual plants, it is suggested that allowances should be considered mainly for the SOFC surface and recuperator surface, assuming values higher, of a certain margin, with respect to the design values. This would allow, at

commissioning level, to mitigate the negative impact of technological uncertainties, allowing to obtain at least the design efficiency and/or design power output. Furthermore, such margins would positively impact on the plant life cycle, because they would alleviate the impact of fuel cell degradation along the projected 5 years life of SOFC stack.



(A) Payback period (B) Internal rate of return

FIGURE 5.23: Monte carlo probability density function outputs - off-design analysis

5.7 Concluding remarks

In this section a method for the analysis of hybrid systems under uncertainties was presented, considering a case study of a turbocharger hybrid system featuring a Solid Oxide Fuel Cell (SOFC) stack. The analysis was divided into three main parts which have been carried out along the Ph.D studies:

- Analysis of the propagation of the uncertainties on the performance of the system
- Design under uncertainty of the system and robust design
- Off-design analysis of the system subject to technological uncertainties

In all cases it was chosen to make use of a response surface metamodel in order to simplify the Matlab-Simulink model from a computational point of view to be able to perform an uncertainty quantification study which is normally computationally expensive. The choice of the response surface technique fell on the Central Composite Face Centered (CCF), and the experiments were carried out with the help of the software Design-Expert.
In the first study, the one related to the analysis of the propagation of uncertainties on the performance of the system, seven different inputs have been considered affected by uncertainty and their effects on power produced, Steam-To-Carbon Ratio (STCR), percentage of mass flow through the wastegate valve and turbocharger rotational speed has been evaluated.
Such parameters were:

- recuperator pressure drop at hot (dp_{hot}) and cold side (dp_{cold})
- anode ejector diameter (d_{ej})
- fuel cell ohmic losses corrective coefficient (k_{ohm})
- gas composition (CO_2 percentage)
- compressor (η_c) and turbine efficiency (η_t).

After thorough evaluation of the response surfaces created in comparison to the simulation model, a study of the impact of such inputs on the aforementioned outputs has been performed.
Results showed that the fuel cell ohmic resistance, the CO_2 percentage in fuel and the ejector diameter are the factors which present the strongest impact on performance, thus they are crucial to be determined accurately, with the lowest standard deviation possible, to guarantee the power and the efficiency desired. On the other hand, pressure drops on the cold and hot side of the recuperator have no significant impact on the considered outputs. Regarding the efficiency of turbine and compressor, they only have impact on the percentage of mass flow which passes through the waste gate valve and a small impact on the rotational speed of the turbocharger, as expected from the design of the system itself. The Response Surface Methodology (RSM) has allowed to evaluate also the interaction effects between the inputs considered; in particular, one of the most important results has been identified in the interaction between the ejector diameter and the CO_2 fuel percentage while considering the STCR as monitored output. In fact, at high values of ejector diameter the STCR is remarkably influenced by the fuel CO_2, while this influence becomes less relevant at low level of the ejector diameter. Regarding the wastegate opening, it can be able to accommodate for the input uncertainties, thus providing effective control of the SOFC temperature, provided that its nominal opening value is high enough to allow different operating conditions, as dictated by the input uncertainties.
Finally, a Monte Carlo simulation was applied to the different response surfaces to evaluate the probability density functions of the outputs. This analysis has allowed evaluating the mean, the standard deviation and the coefficient of variance of the monitored outputs. In particular, by comparing the coefficient of variance (COV) of the outputs to the input ones, it has been possible to highlight that the model, hence the plant, is able to mitigate the propagation of uncertainties, as the COV of the outputs is lower than the COV of the inputs. This result is particularly interesting as it means that, for this case study, the net power production would be more stable and close to the nominal value despite the input uncertainties.
Starting from the results obtained in this first part of the study, it was decided to analyze the system from a design point of view, to evaluate how would the sistem perform with the replacement of the SOFC stack and of the recuperator. The area of the solid oxide fuel cell (A_{cell}), the current density (J_{cell}) and the recuperator surface (A_{REC}) have been considered as factors to build the response surfaces. The effect of uncertainties in the turbocharger maps in terms of turbine and compressor efficiency (η_t, η_c) and in the ohmic losses of the SOFC stack (k_{ohm}) were also considered when building the response surfaces, evaluating the model 15 times per point and assuming a normally distributed value for the stochastic parameters.
The response surfaces were built to evaluate the performance of the system, in terms of net power and net efficiency, and its economic profitability, in terms of capital costs, payback period and internal rate of return. Three different market scenarios were considered regarding electricity and natural gas costs: Italy, United Kingdom and United States (NY).

The uncertainties had an impact on the regression parameter R^2 of the response surfaces, as for each point the solutions were spread and so it was not possible to fit a surface accurately. However, the R^2 was generally high, with values higher than 0.9 and in most cases higher than 0.95. The response surfaces highlighted that the net efficiency and net power depended mostly on the cell area and stack current density, as the power produced is related to the SOFC stack performance only. In particular, the highest net power production would be achieved using a larger stack (high A_{cell} and high J_{cell}), while the net efficiency presented a maximum for low J_{cell}, as this would impact directly the ohmic losses and then the efficiency.
From a first-order economic analysis it was highlighted that United Kingdom presented the lowest payback period and internal rate of return, followed by New York and then by Italy. This is related to the fact that Italy presents a very high cost of natural gas, which is almost double the one of United Kingdom and New York. The electricity cost instead was evaluated as a saving that the plant owner would earn from self-consuming the power of the fuel cell hybrid system. This formulation was then beneficial for United Kingdom as it presents quite high costs of electricity, while less beneficial for New York which has the lowest electricity price. Analysing the economic responses, it was observed that the payback period was mostly influenced from the current density, presenting a minimum for high values of J_{cell} and A_{cell}, corresponding to a high net power production. A similar behaviour could also be highlighted for the internal rate of return, as expected.
The polynomial response surfaces of net power and capital costs were then used to perform a multi-objective optimization with genetic algorithm and a Pareto front was obtained. The Pareto front allowed to evaluate the optimum solutions in terms of costs and power production, pointing out the two "top power production" solutions. These solutions in fact presented the highest net power (35.6kW and 36 kW), with a very similar economic investment in terms of capital costs (54.4 k€and 54.8k€). A robust design analysis of these solutions was then carried out, building new response surfaces with fixed design parameters (A_{cell}, J_{cell}, A_{rec}) and considering instead as factors the uncertain parameters (η_t, η_c, k_{ohm}). A Monte Carlo simulation was performed on such response surfaces to evaluate the coefficient of variance and so the robustness of the solutions. This analysis showed that the point with the highest net power production (36kW) presented a lower coefficient of variance both when considering power and both when considering efficiency, resulting in a more robust solution. This result is particularly interesting from the point of view of the manufacturer, as it would mean that choosing this solution, despite a slightly higher capital cost, would guarantee less uncertainties towards customers, a more reliable design production less affected by uncertainties and tolerances of the system.
Finally, the latter study focused on the off-design analysis of the system, starting from the solution with the highest internal rate of return considering the Italian market scenario from the previous work as the design point. An off-design analysis of such design point was then performed considering three different loads: 100%, 85% and 70% electrical power output. The power load was modulated acting on the electrical current at constant utilization factor (i.e. fuel flow reduced proportionally). The response surfaces of net power and net efficiency were then built for each load percentage, using the uncertainties as factors to evaluate their impact on system performance. From this analysis it was highlighted that one of the points of 100% load condition, estimated by the metamodel, was not feasible, as the uncertainties were contributing to raise the SOFC stack temperature over the threshold. This particular behaviour led to a lower R^2 for the responses at 100%, (about 0.81), while the R^2 of net power and net efficiency at 85% and 70% were very close to 1.0, meaning that

the surfaces were approximating the model behaviour almost perfectly.
The polynomial responses were then used to perform a Montecarlo simulation of the system, assuming that the system would work for one third of the total operating hours per year at each different load. The economic uncertainties related to fuel and electricity prices were considered at this level, assuming no particular correlation amongst them. This analysis allowed to evaluate the probability density function of the revenues of the plant, in particular the internal rate of return and the payback period. The resulting probability density function showed that, as expected, the variability of the prices of fuel and electricity has a strong impact on both IRR and PBP, leading to a coefficient of variance of about 10% for both parameters. The results suggested that, to mitigate the impact of technological uncertainties, allowances should be considered at the design phase mainly for the SOFC surface and recuperator surface, as well as a proper evaluation of the prices along the years. This would allow to reach the design efficiency and/or design power output at commissioning, alleviating at the same time the impact of fuel cell degradation along the projected life of SOFC stack.
Overall a solid and working method for the analysis of such systems subject to technological uncertainties was presented, showing that the application of uncertainty quantification methods can be crucial when more insights are needed both from a performance and from an economic point of view.

5.8 Annex A

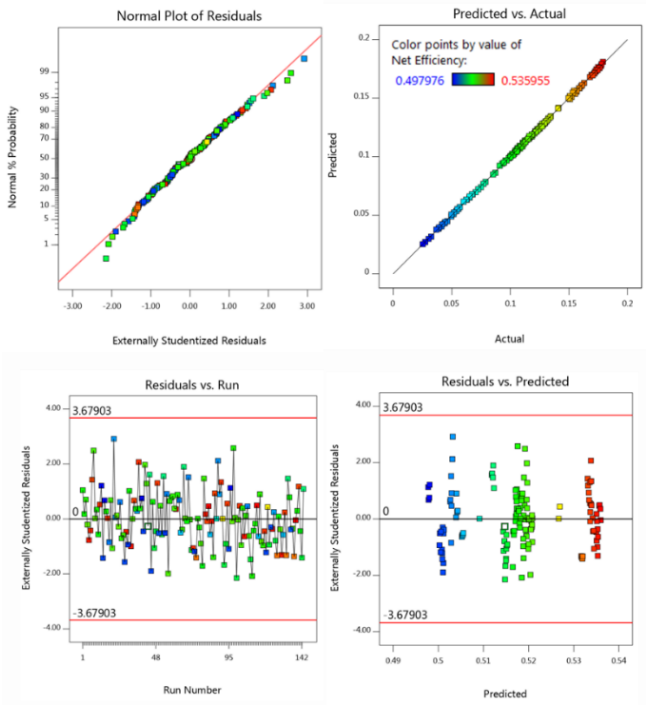


FIGURE 5.24: Diagnostic plots related to net power - UQ analysis on hybrid system

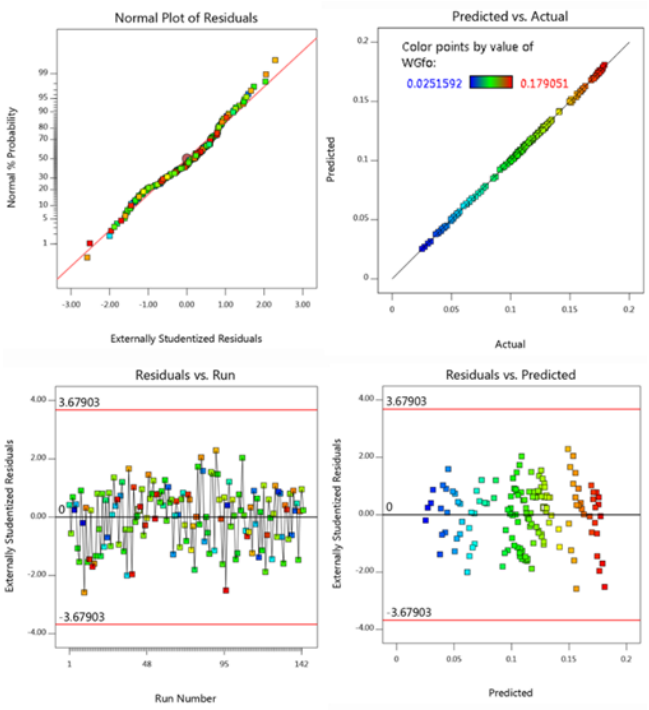


FIGURE 5.25: Diagnostic plots related to $\dot{m}_{WG}$ - UQ analysis on hybrid system

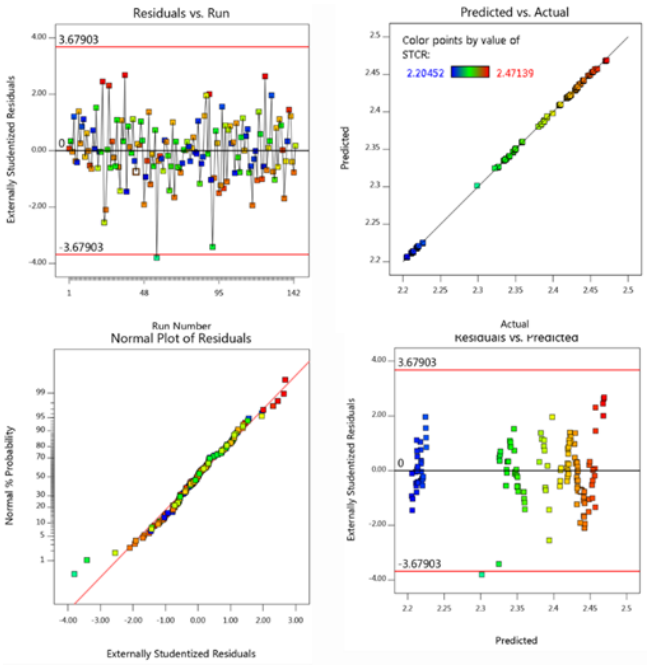


FIGURE 5.26: Diagnostic plots related to STCR - UQ analysis on hybrid system

References

[18] R. Ghanem, H. Owhadi, and D. Higdon. *Handbook of uncertainty quantification*. Ed. by Springer. Berlin, Germany, 2017, pp. 1–2053. ISBN: 9783319123851. DOI: 10.1007/978-3-319-12385-1.

[23] A. Cuneo, A. Traverso, and S. Shahpar. "Comparative analysis of methodologies for uncertainty propagation and quantification". In: *Proceedings of the ASME Turbo Expo*. Vol. 2C-2017. American Society of Mechanical Engineers, 2017. ISBN: 9780791850800. DOI: 10.1115/GT2017-63238. URL: https://asmedigitalcollection.asme.org/GT/proceedings/GT2017/50800/Charlotte,NorthCarolina,USA/241782.

[40] A. Giugno, A. Cuneo, and A. Traverso. "Analysis of uncertainties in compact plate-fin recuperators for microturbines". In: *Applied Thermal Engineering* 150.September 2018 (2019), pp. 1243–1251. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2019.01.093. URL: https://doi.org/10.1016/j.applthermaleng.2019.01.093.

[42] A. Giugno, L. Mantelli, A. Cuneo, and A. Traverso. "Performance analysis of a fuel cell hybrid system subject to technological uncertainties". In: *Applied Energy* 279 (2020). ISSN: 03062619. DOI: 10.1016/j.apenergy.2020.115785.

[61] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. 2009. ISBN: 9780387781648. arXiv: arXiv:1011.1669v3.

[74] L. Cassettari, R. Mosca, and R. Revetria. "Monte Carlo simulation models evolving in replicated runs: A methodology to choose the optimal experimental sample size". In: *Mathematical Problems in Engineering* 2012 (2012), pp. 73–76. ISSN: 1024123X. DOI: 10.1155/2012/463873.

[123] L. Galanti and A. F. Massardo. "Micro gas turbine thermodynamic and economic analysis up to 500 kWe size". In: *Applied Energy* 88.12 (2011), pp. 4795–4802. ISSN: 03062619. DOI: 10.1016/j.apenergy.2011.06.022.

[137] X. Zhang, S. H. Chan, G. Li, H. K. Ho, J. Li, and Z. Feng. "A review of integration strategies for solid oxide fuel cells". In: *Journal of Power Sources* 195.3 (2010), pp. 685–702. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2009.07.045.

[138] A. B. Stambouli and E. Traversa. "Solid oxide fuel cells (SOFCs): A review of an environmentally clean and efficient source of energy". In: *Renewable and Sustainable Energy Reviews* 6.5 (2002), pp. 433–455. ISSN: 13640321. DOI: 10.1016/S1364-0321(02)00014-X.

[139] V. Mehta and J. S. Cooper. "Review and analysis of PEM fuel cell design and manufacturing". In: *Journal of Power Sources* 114.1 (2003), pp. 32–53. ISSN: 03787753. DOI: 10.1016/S0378-7753(02)00542-6.

[140] M. A. Azizi and J. Brouwer. "Progress in solid oxide fuel cell-gas turbine hybrid power systems: System design and analysis, transient operation, controls and optimization". In: *Applied Energy* 215 (2018), pp. 237–289. ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.01.098.

[141] A. Cuneo, V. Zaccaria, D. Tucker, and A. Sorce. "Gas turbine size optimization in a hybrid system considering SOFC degradation". In: *Applied Energy* 230 (2018), pp. 855–864. ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.09.027.

[142] M. Henke, J. Kallo, K. A. Friedrich, and W. G. Bessler. "Influence of pressurisation on SOFC performance and durability: A theoretical study". In: *Fuel Cells*. 2011. DOI: 10.1002/fuce.201000098.

[143] K. Venkataraman, E. C. Wanat, and L. D. Schmidt. "Steam reforming of methane and water-gas shift in catalytic wall reactors". In: *AIChE Journal* 49.5 (2003), pp. 1277–1284. ISSN: 00011541. DOI: 10.1002/aic.690490518.

[144] STAXERA. *SOFC Stack - Mk200, Product Data Sheet*. Tech. rep.

[145] L. Mantelli, M. L. Ferrari, and L. Magistri. "Off-design performance analysis of a turbocharged solid oxide fuel cell system". In: *Applied Thermal Engineering* 183.P1 (2021), pp. 116–134. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2020.116134. URL: https://doi.org/10.1016/j.applthermaleng.2020.116134.

[146] A. Traverso, A. F. Massardo, and R. Scarpellini. "Externally Fired micro-Gas Turbine: Modelling and experimental performance". In: *Applied Thermal Engineering* 26.16 (2006), pp. 1935–1941. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2006.01.013.

[147] M. L. Ferrari, M. Pascenti, and A. F. Massardo. "Ejector Model for High Temperature Fuel Cell Hybrid Systems: Experimental Validation at Steady-State and Dynamic Conditions". In: *Journal of Fuel Cell Science and Technology* 5.4 (2008), pp. 041005–1–7. ISSN: 1550624X. DOI: 10.1115/1.2890102.

[148] A. Greco, A. Sorce, R. Littwin, P. Costamagna, and L. Magistri. "Reformer faults in SOFC systems: Experimental and modeling analysis, and simulated fault maps". In: *International Journal of Hydrogen Energy* 39.36 (2014), pp. 21700–21713. ISSN: 03603199. DOI: 10.1016/j.ijhydene.2014.09.063.

[149] A. Sorce, A. Greco, L. Magistri, and P. Costamagna. "FDI oriented modeling of an experimental SOFC system, model validation and simulation of faulty states". In: *Applied Energy* 136 (2014), pp. 894–908. ISSN: 03062619. DOI: 10.1016/j.apenergy.2014.03.074.

[150] M. L. Ferrari. "Advanced control approach for hybrid systems based on solid oxide fuel cells". In: *Applied Energy* 145 (2015), pp. 364–373. ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.02.059.

[151] R. L. Wasserstein and N. A. Lazar. *The ASA's Statement on p-Values: Context, Process, and Purpose*. 2016. DOI: 10.1080/00031305.2016.1154108.

[152] A. Cuneo, V. Zaccaria, D. Tucker, and A. Traverso. "Probabilistic analysis of a fuel cell degradation model for solid oxide fuel cell and gas turbine hybrid systems". In: *Energy* 141 (2017), pp. 2277–2287. ISSN: 03605442. DOI: 10.1016/j.energy.2017.12.002.

[153] O. Harel. "The estimation of R2 and adjusted R2 in incomplete data sets using multiple imputation". In: *Journal of Applied Statistics* 36.10 (2009), pp. 1109–1118. ISSN: 02664763. DOI: 10.1080/02664760802553000.

[154] A. Giugno, L. Mantelli, and A. Traverso. "Robust Design of a Fuel Cell-Turbocharger Hybrid Energy System". In: *Proceedings of the ASME Turbo Expo*. 2020.

[155] A. Arsalis. "Thermoeconomic modeling and parametric study of hybrid SOFC-gas turbine-steam turbine power plants ranging from 1.5 to 10 MWe". In: *Journal of Power Sources* 181.2 (2008), pp. 313–326. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2007.11.104.

[156] A. Giugno, L. Mantelli, A. Cuneo, and A. Traverso. "Robust Design of a Hybrid Energy System". In: *E3S Web of Conferences*. Vol. 113. 2019. DOI: 10.1051/e3sconf/201911302008.

[157] Chemical Engineering. *CEPCI Archives - Chemical Engineering*. URL: https://www.chemengonline.com/tag/cepci/ (visited on 01/15/2021).

[158] Eurostat - Statistical Office of the European Communities. *Electricity prices components for non-household consumers - annual data*. Tech. rep. 2019. URL: https://ec.europa.eu/eurostat/data/database.

[159] EIA - U.S. Energy Information Administration. *Average Price of Electricity to Ultimate Customers by End-Use Sector*. Tech. rep. 2019. URL: https://www.eia.gov/electricity/monthly/epm{_}table{_}grapher.php?t=epmt{_}5{_}6{_}a.

[160] Eurostat - Statistical Office of the European Communities. *Gas prices components for non-household consumers - annual data*. Tech. rep. 2019. URL: https://ec.europa.eu/eurostat/data/database.

[161] EIA - U.S. Energy Information Administration. *Natural Gas Prices*. Tech. rep. 2019. URL: https://www.eia.gov/dnav/ng/ng{_}pri{_}sum{_}dcu{_}SNY{_}a.htm.

[162] Statease. *Interpretation of R-squared*. URL: https://www.statease.com/docs/v11/contents/analysis/interpretation-of-r-squared/ (visited on 01/15/2021).

[163] T. Murata and H. Ishibuchi. "MOGA: multi-objective genetic algorithms". In: *Proceedings of the IEEE Conference on Evolutionary Computation*. Vol. 1. 1995, pp. 289–294. DOI: 10.1109/icec.1995.489161.

[164] C. M. Fonseca and P. J. Fleming. "Genetic Algorithms for Multiobjective Optimization :" in: *Icga* July (1993). ISSN: 14639076. DOI: citeulike-article-id:2361311.

Chapter 6

Thermo-economic assessment of a combined cycle under market uncertainties

6.1 Introduction

The challenging targets of Green House Gas (GHG) emission reduction, that have been set by the European Union (EU) and other international institutions [165], has led the EU power generation mix to change considerably in favor of renewables, which have become the fastest growing source of electricity among the member countries [166]. However, non-dispatchable Renewable Energy Sources (RES), e.g. wind and solar, are characterized by an extreme hour by hour production variability, due to their stochastic and non-programmable nature. Variability in production is reflected in electricity prices and this particularly affects natural gas-fired Combined Cycle power plants, which are currently the backbone of EU electrical grid and are foreseen by the EU as the bridging technology (till the horizon of 2050) to a decarbonized scenario, thanks to their reduced carbon footprint and fast response in terms of grid stabilization [167]. Future energy strategies may be based on alternative fuel (i.e. clean hydrogen or ammonia) [168, 169] but in the short-term strategies the role of Combined Cycle Gas-Turbine (CCGT) in the electricity market scenario is considered to be essential, having therefore shifted to a backup capacity for grid stabilization and to a spinning reserve operation, making flexibility the key feature, also supported by energy storage technologies [170].
The mothballing or the closure of less flexible, and therefore less profitable, plants occurred in past years, affecting the installed capacity of Power Oriented CC (i.e. devoted to electrical energy production only) [171]. In fact, at EU level at the end of 2013, 24.7 GW, about 12.8%, of gas-fired power capacity were idled, closed or at risk of closure [172]. Some gas-fired assets were able to survive thanks to an enhanced flexibility, which can be exploited to provide ancillary services. Flexibility is often pursued by Combined Heat and Power (CHP) or tri-generative arrangement. CHP CCGTs were mainly installed to satisfy a local low temperature thermal demand via District Heating Network (DHN), or close to industrial sites to provide high temperature process steam, so are usually maintained operative even if less profitable in the electricity market, creating a useful back-up capacity for grid resilience. This

capacity can be strengthened by installation of new CHP CCGT if the economic conditions are sustainable.
District Heating Networks may also constitute an interesting option to limit GHG, since centralized heat generation, using larger combustion units with higher energy efficiency and more advanced control over air pollution, has been considered as an efficient, environmentally friendly and cost-effective method for heating purposes in many high-density city buildings [173]. The core element of a DHN can be a Heat-Only Boiler (HOB), a CHP plant, but also a large Heat Pump (HP), often considered as a smart way to provide clean energy for heating purposes via DHNs [174]. Usually DHNs rely on a combination of several generators together. In particular, cogeneration plants (e.g. Combined Cycle, Internal Combustion Engine and Steam power plant) are usually adopted to cover the base of the thermal load while HOBs cover demand peaks or supply back-up generation. When a heat source is available at a temperature lower than the DHN temperature level, large-scale heat pumps can be introduced to enhance its temperature. The advantage of using this technology is its efficiency (e.g., compared to electrical boilers) and its ability to use inexpensive sources of heat to balance the DHN as a heat generator and the electricity grid as an electrical load. Seven types of heat sources were identified for the large-scale HPs operating in connection with a DHN. Ordered by installed capacity there are: sewage water, ambient water, industrial waste heat, geothermal water, flue gas, District Cooling return and solar heat storages [175]. Flue gas-based installations, thanks to a higher temperature heat source can have better performance and can be coupled to the main DHN CHP generators.
The potentialities of using a combination of CHP plants and large scale heat pumps in district heating systems to balance intermittent renewable power production has been discussed by Levihn [176], presenting lessons and empirical data from operating such a system in Stockholm DHN. Blarke and Dotzauer [177] also described the integration of a heat pump and a natural gas-fired CHP plant, demonstrating the enhancement of the global fuel to energy efficiency within a context of electricity price variability. Another study by Ommen investigated several layout solutions to optimally couple heat pumps to CCGTs working in CHP mode [178]. It is well known that a good heat source, in terms of temperature level, stability and availability of heat, strongly enhances heat pump performance. Flue gas condensation has been considered an interesting option by many researchers, since it presents quite a high and constant temperature level and also allows one to exploit the latent heat potential of the flue gases, increasing beyond the unit the global efficiency of the power plant evaluated with respect to the fuel Lower Heating Value (LHV) [179, 180, 181]. The cost of investment in such technology could be very high [182], and a careful design is necessary to overcome the risk of not paying back the investment. Some economic conclusions have been drawn regarding profitability sensitivity to a large number of parameters (e.g. electricity costs, investment costs, gas costs and heat exchanger design [183, 179]), but, to the best of the authors' knowledge, there is no study in the literature considering the uncertainty of forecast of these parameters in this kind of analysis.
This work aims to assess the economic viability of a CHP CCGT coupled with a flue gas condensing HP or with a HOB, considering possible future electricity and market scenarios and the uncertainties related to prices of gas and electricity to evaluate their impact on such system. This work was developed within the framework of an H2020 EU project called PUMP-HEAT [184].

6.2 Methodology

The design under uncertainty of a CCGT coupled with a large-size condensing HP, was performed considering as benchmark a CCGT coupled with a HOB. To evaluate the economic viability of the system within a 20-year span (2020-2040), the variability of the future prices forecast should be considered, as there is no certainty about the values that they will take on in coming years. A case study based on the Italian market is proposed but the method presented has a general application.
The expected distribution of the single national price of electricity (PUN) was modeled with probability density functions trained on historical data, building an hourly profile of the PUN which would take into consideration the uncertainties derived from market variability. A similar process was done for the gas cost, building however a single probability density function with a standard deviation which could take into account the uncertainties about the forecasted values. To account for the uncertainties, a Monte Carlo simulation was performed, adopting a number of samples which could minimize the mean square pure error (MSPE) [77], on a deterministic model built to simulate the behavior of the CCGT-HP and CCGT-HOB systems, subject to a deterministic daily thermal demand. The economic Key Performance Indicator (KPI) can be assessed on a yearly basis, not just as deterministic values but including the effect of input uncertainties and thus the risk of a wrong forecast. In Figure 6.1, a simplified scheme of the method described above is presented to clarify the process.
In Section 6.3 the description of the deterministic thermodynamics model is carried

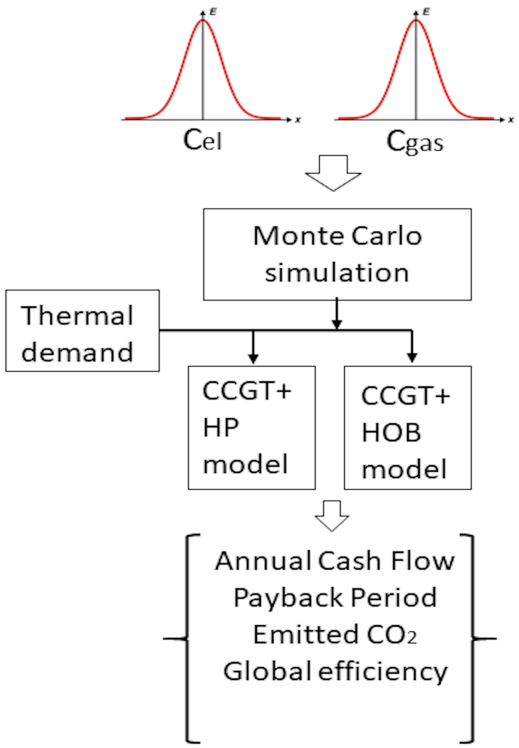


FIGURE 6.1: Scheme of the methodology for the design under uncertainty of the CCGT-HP/CCGT-HOB system

out, while the assumptions related to the capital expenditure, the operating costs and the modeling of the gas cost, price of heat and electricity price are presented in the third section. The results are then plotted and discussed in Section 6.5.

6.3 Model description

The case study is based on a standard 400 MW three-pressure-level CCGT CHP power plant, based on an F Series gas turbine, feeding a second generation DHN with 120°C and 70°C of forward and return temperature respectively [185]. The heat generation is guaranteed by steam extraction upstream from the Low Pressure (LP) turbine. The extracted steam is then redirected to a heat exchanger (main DHN-HX) which delivers the heat to the user. The steam extraction is followed by a decrease in the mass flow rate through the LP turbine and consequently in power generation. The Combined Cycle model was developed using Gate Cycle [186], and validated against available public data for thermal production. The complete Iron Diagram (i.e. the CHP operating range represented on an electrical power vs thermal power graph) was characterized through 42 complete calculations increasing the thermal load up to the maximum steam extraction level (blue continuous line in Figure 6.2) at different relative GT percentage loads (100%, 75%, 60%, 45%), at ISO ambient condition (i.e. 15°C, 101325 Pa, 60% Relative Humidity) with the water condenser cooling flux at 12°C. The CCGT global efficiency (eq. 6.1), solid line, grows with the increase of heat production up to a maximum at 89.5% at 261 MWth for the 100% GT load, while the electrical efficiency decreases with steam extraction and GT load.

Two systems are proposed to integrate the CCGT to increase its operative thermal

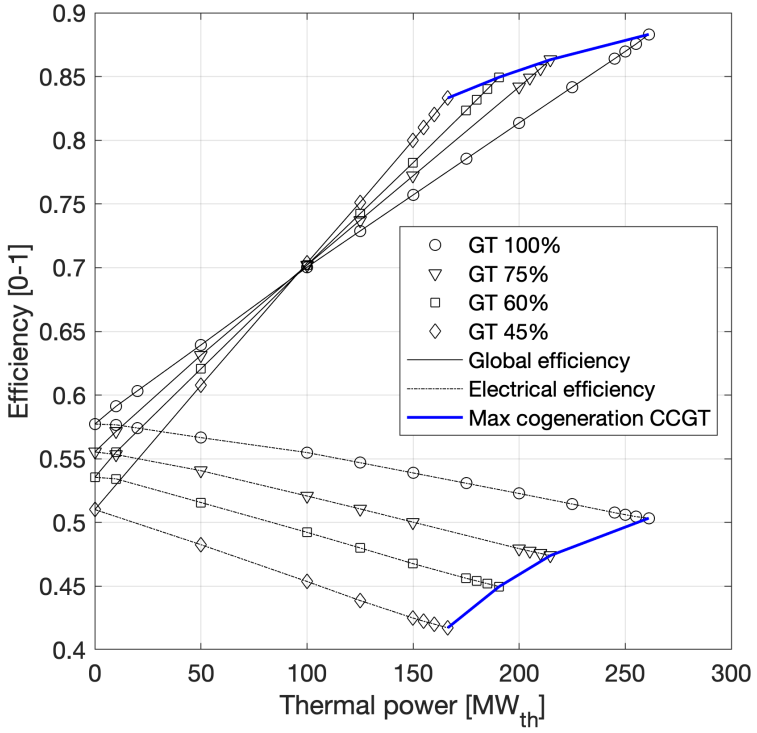


FIGURE 6.2: CHP cycle efficiency

range:

1. a flue gas condensing heat pump
2. a Heat-Only Boiler (HOB).

The global efficiency was used to compare the cogenerative performance, calculated as:

$$\eta_{gl} = \frac{P_{el} + P_{th}}{L\dot{H}V} \tag{6.1}$$

Where P_{th} and P_{el} are the thermal and electrical net power outputs of the plant and $L\dot{H}V$ is the primary energy consumption evaluated over the Lower Heating Value of the Natural Gas.

6.3.1 Heat-Pump integration

The solution integrating the flue gas condensing heat pump aims to recover both the sensible heat and the latent heat from the flue gases, significantly improving the global efficiency of the plant.
Nevertheless, the exploitation of this energy potential implies some technological problems, imposing a particular layout configuration in order to avoid high costs and efficiently perform the heat recovery process. The latent heat can be exploited only if the flue gases are cooled below the dew point (50-55°C depending on the chemical composition) and condensation occurs. On the other hand, a minimum temperature level (60°C) at the stack is required in order to guarantee a sufficient buoyancy effect to increase pollutant dispersion. Consequently, a reheating of flue gas must be considered after the latent heat exploitation. To avoid a huge gas-gas regenerative surface, thus high costs, an indirect reheating performed by the DHN returning water was preferred. The condensing flue gases were selected as the heat pump's low temperature heat source with a twofold function: upgrading the temperature level of the latent heat recovered below the dew point and using the electricity generated by the CCGT to satisfy the heat demand with enhanced flexibility and possible economic advantages during low electrical price period.
It has been demonstrated that it is preferable to link the heat pump in series with the Heat Exchanger (HX) that transfers the heat from the CCGT bottoming cycle to the DHN, the DHN-HX. These considerations and constraints led to the plant layout shown in Figure 6.3 where a scheme of the integrated solution CCGT with a flue gas condensing heat pump is shown. The heat pump system is designed to have 30 MWel power, working in the best conditions with a Coefficient Of Performance (COP) of 4.55 at maximum load, and a thermal capacity of 137 MWth. A set of five 6 MWel Heat Pumps working in parallel were used to reach this heat production capacity in order to reduce the severity of off-design operations and to match available equipment size [187].
Nevertheless, part of the heat that the heat pump extracted from the Flue Gas (FG) condenser has to be provided to the flue gas in the reheater in order to guarantee 60°C at the stack. The heat exchanged in the reheater by the DHN return water is restored by the HP, hence with the heat pump working at maximum load the increase in final power plant thermal output is about 110 MWth.

A Matlab model, combining the CCGT calculations carried out with the Gate Cy-

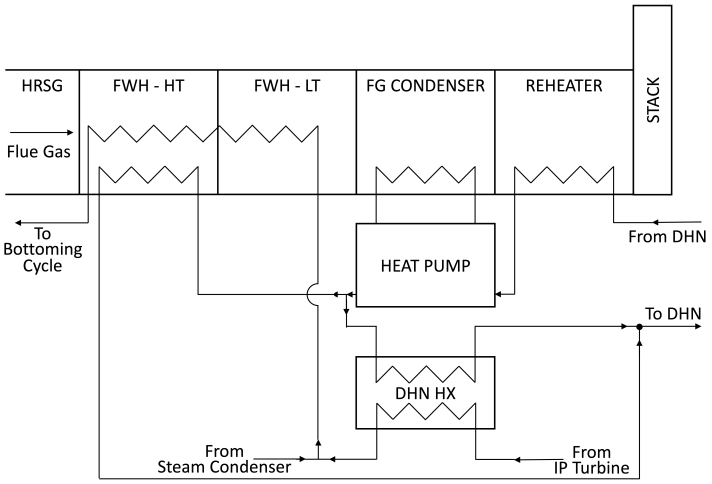


FIGURE 6.3: CCGT HP layout scheme

cle model, was developed to assess the integrated layout performance. The model includes the energy and mass balances of the components added between the Heat Recovery Steam Generator (HRSG) and the stack, the psychometrics of flue gas and the model of the heat pumps. The heat balance of the integrated layout was solved thanks to a nested iterative logic schematized as in Figure 6.4.

Firstly, the inner cycle iterates on the temperature of water at the heat pump outlet that enters the Feed Water Heating High Temperature (FWH-HT) and the DHN-HX. This value is used to calculate the heat pump's COP and so the heat transferred from the condensing flue gas to the water exiting the flue gas reheater. A second level of iteration is performed on the total DHN mass flow, thus on the overall thermal production. The new value for this iteration is computed by summing the amount of mass flow which can be heated up to 120°C from the FWH-HT and the DHN-HX respectively. Finally, the last iteration level is on the temperature entering the Feed Water Heating Low Temperature (FWH-LT) as a result of the mixing process of steam coming from the DHN-HX and the main condenser of CCGT bottoming cycle. The heat pump model considered in this study is a standard compression vapor heat pump including a regenerator, with butane (R600) as working fluid. The cycle is modeled to maximize, with the given source and sink temperature, the COP within the respect of some constraints mainly related to minimum temperature differences required at the heat exchangers. The whole cycle and the thermodynamic properties were modeled and computed using the CoolProp database [188]. The main assumptions are reported in Table 6.1 and are maintained constant also during off-design operation. So the calculation of the COP is mainly related to the temperature level available in the flue gases at the different GT load.

To accurately model the flue gas condensation process, the psychometrics equa-

TABLE 6.1: Heat pump modeling assumptions

Fluid	ΔT_{SH}	η_c	ΔT_{eva}	ΔT_{cond}	ΔT_{reg}
R600	25 °C	0.64	3°C	4°C	3°C

tions reported in the ASHRAE Handbook [189] were taken as reference, substituting

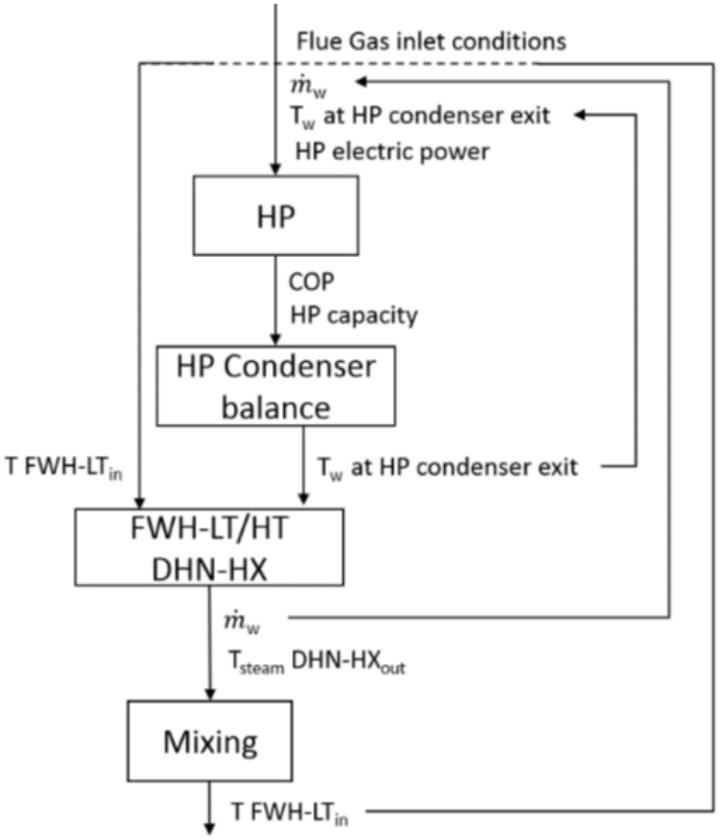


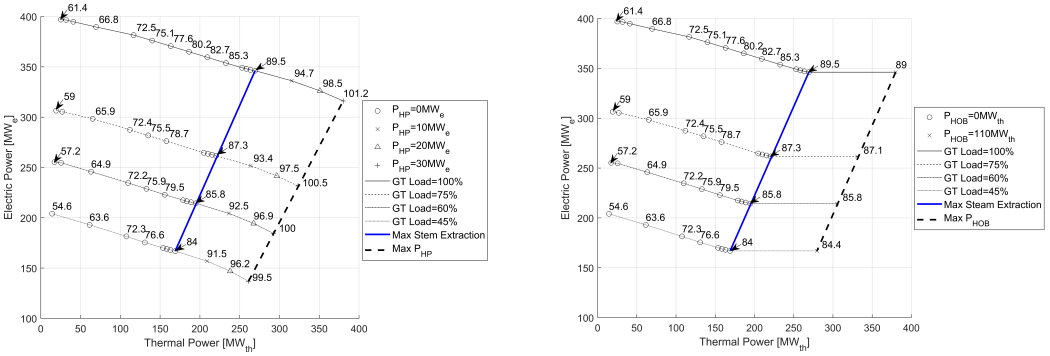
FIGURE 6.4: CCGT model iterative logic

the specific heat of dry air with the dry-flue-gas value.
A set of 5 Heat Pumps working in parallel (equally sharing the load between the activated machines) were used to extend the heat production capacity of the integrated system, over the maximum CCGT cogeneration line, blue line, as shown in Figure 6.5A. The label presents the value of the global efficiency that can be further enhanced by the HP up to 101.2% at 371 MWth. Theoretically it would also be possible to use the heat pumps in partial steam extraction conditions, but the efficiency advantages are too modest to justify such an operational complication.

6.3.2 Integration with the heat-only boiler

The integrated solution described in the previous section (CCGT-HP) is compared with an integration of the same CCGT power plant with a HOB (CCGT-HOB), a gas-fired device coupled with the CHP plant to cover the heat peaks that exceed the CCGT maximum thermal output. The boiler was sized to provide the same thermal output increase of the heat pump, 110 MWth. The boiler energy efficiency was assumed to vary with the load, according to the Satyavada and Baldi model [190], therefore linearly decreasing between 0.875 at 5% load, the minimum value, and 0.86 at design point. Condensation of flue gases within the boiler is inhibited since the DHN cold water return temperature, 70°C is higher than condensation temperature. Figure 6.5B shows the diagram of the possible operating conditions of the CCGT-HOB solution. It is important to remark that, while in the CCGT-HP Iron diagram

(Figure 6.5A) the iso-GTload lines are also iso-fuel lines, this is not true once the HOB is activated, since it burns additional fuel to cover the heat demand. However, as shown by Figure 6.5B labels, activating the HOB the global efficiency of the integrated CHP plant is not affected significantly, since the boiler efficiency is always comparable to the reference CHP efficiency in full extraction conditions. The CCGT-HOB solution increases the thermal production without a further reduction of the electrical output; this can be beneficial to the system profitability under high electricity price conditions. On the contrary, the CCGT-HP allows provision of the same additional heat without increasing fuel consumption. Thus, it is expected to save primary energy and reduce emissions, while from the economic point of view, it is expected to be more viable when electricity prices are low.



(A) CCGT-HP (B) CCGT-HOB

FIGURE 6.5: Iron diagrams (global efficiency shown in labels)

6.4 Economic assumptions

The market conditions during the plant lifetime are extremely relevant for the economic viability of the investment in a new CHP plant. In this section, the assumptions made to assess the CCGT-HP and CCGT-HOB solutions are described. The lifetime of such power plants is expected to be of twenty years. Thus, all the assumptions are indicatively referred to the period 2020-2040.

6.4.1 Capital expenditure and operation and maintenance costs

To assess the economic investment and operating costs, the data reported by the Technology Data catalogue of the Danish Energy Agency have been used as a reference [182]. The predicted CAPital EXpenditure (CAPEX) for 2020 and the expected Operation and Maintenance (O& M) costs in the plant's lifetime for the CCGT, the heat pump, and the HOB are reported in the Table 6.2.

With respect to the reference CAPEX, based on a 4MWth heat pump, to fulfill the requirements of 137 MWth design thermal output, a set of 5 HPs with a rated output of 27.4 MWth was selected. Such large-scale heat pumps require 6 MWel each, with an increase in size of about 6 times for the electrical driven centrifugal compressor

TABLE 6.2: Heat pump modeling assumptions

	CCGT	HP	HOB
CAPEX	0.88 M€/MWel	0.66 M€/MWth	0.06 M€/MWth
O& M fixed	27800 €/MWel/yr	2000 €/MWth/yr	1900 €/MWth/yr
O& M variable	4.2 €/MWhel	1.7 €/MWhth	1.0 €/MWhth
Size	400 MWel	137 MWth	110 MWth

and of about 6.85 times the Heat Exchanger size. Using the cost function provided by [191], a reduction of the specific purchasing cost of 9.7% for the compressor and 12.9% for the Heat Exchanger was estimated, so an average cost reduction of 10% of the expected specific capital expenditure was introduced in the HP CAPEX calculation. The resulting costs for the proposed CCGT-HP solution are 433.12 M€of CAPEX, 11.39 M€per year of fixed O& M and 4.2 €/MWhel for the CCGT plus 1.7 €/MWhth for the heat pump of variable O& M. While for the CCGT-HOB the CAPEX is of 358.64 M€, 11.33 M€per year of fixed O& M and 4.2 €/MWhel for the CCGT plus 1.0 €€/MWhth for the HOB of variable O& M.

6.4.2 Gas and heat price

As concerns the gas cost, an average forecasted trend was considered, based on the EU Reference Scenario [192]. From Figure 6.6 it can be observed that to take into account the variability of the gas cost in the future, a confidence interval was considered (blue area) around the average projection (line with markers), the uncertainty increases the more the projection is a long-term forecast. The cost of gas was considered normally distributed, averaged on the intermediate projected price and within a probability of 95.5% of being in the confidence interval, as can be observed from the Gaussian distribution plotted in red in Figure 6.6 over the auxiliary axes. The related Gaussian probability density function has a mean of 30 €/MWh and a standard deviation equal to 5 €/MWh.

According to some investigations and statistical reports on the Italian case study, such as [193, 194], the reference price of heat was set to 90 €/MWh. Since a fundamental component of the Levelized Cost of Heating is the cost of the fuel [195], it was supposed that the price of heat would increase, or decrease, as the cost of natural gas, according to the following equation:

$$p_{heat} = p_{heat,ref} + \left(C_{gas} - C_{gas,2018}\right) \tag{6.2}$$

where C_{gas} is the gas cost, determined as described above, $C_{gas,2018}$ is the reference cost of gas (2018), set to 22.83 €/MWh [196] and $p_{heat,ref}$ is the reference price of heat assumed equal to 90 €/MWh. Moreover, the cost for each emitted ton of carbon dioxide was considered, according to the EU-ETS policy. The carbon prices are projected to increase almost linearly from 20 €/ton in 2020 to 60 €/ton in 2040 [192]. Thus, an average value of 40 €/ton can be considered during the plant's lifetime. An emission factor of 55.94 tons per TJ of fuel energetic potential was selected according to Italian government data [197].

The annual thermal demand was modeled on the basis of a scale down of the Turin DHN historical data (2016). The yearly thermal demand of 687 GWh, is assumed to

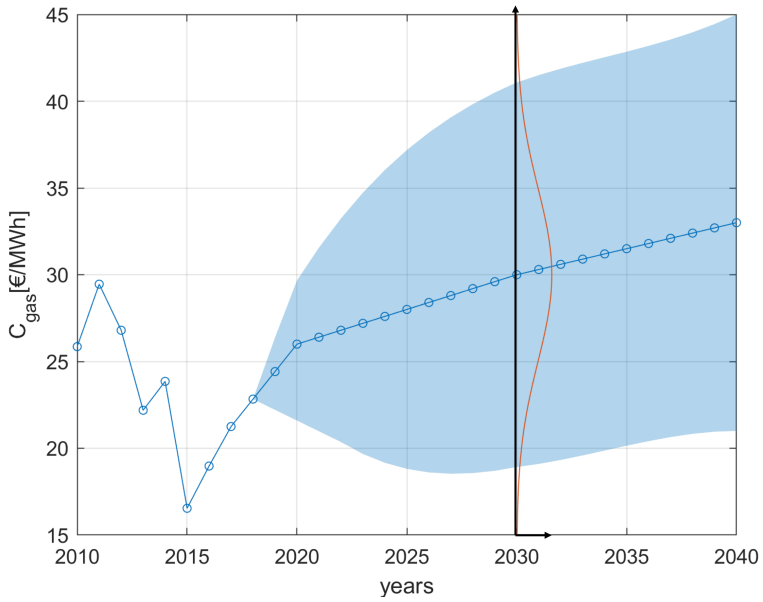


FIGURE 6.6: Historical and projected cost of gas [192]

be concentrated between 15th October and 14th April of each winter season, equal to 182 days. Each daily thermal demand is assumed to be characterized by three main zones: an off-peak period between 10 p.m. and 6 a.m., a peak period between 6 a.m. and 9 a.m. and the medium load period. During these hours the heat demand is assumed to be 1.728, 0.385 and 1.211 times the average demand of the day respectively. The modelling of the standard day heat demand was performed by searching for the periods' duration and the respective heat demand values, minimizing the difference compared to the average day profile obtained from the real data.
The deterministic model of the thermal load demand takes into account standard, warm and cold days. During warm and cold days, the heat demand is scaled by a factor with respect to standard days. The number of both warm and cold days, 76 and 39, and the value of the respective scaling factor, 1.229 and 0.602, were chosen in order to best fit the real data seasonal power profile. The fitted seasonal profile is represented in Figure 6.7, where are also reported the thermal output level of the plant, at minimum and maximum load, with and without integration, while a comprehensive report of the modeled heat demand is provided in Table 6.3.

TABLE 6.3: Heat demand profile data

	Standard days	**Warm days**	**Cold days**
6 a.m. - 9 a.m.	307.97 MWth	185.51 MWth	378.54 MWth
9a.m. - 10p.m.	215.76 MWth	129.97 MWth	265.19 MWth
10p.m. - 6a.m.	68.61 MWth	41.33 MWth	84.33 MWth

6.4.3 Electricity price

The profitability of a power system along its life is highly dependent on the distribution of the electricity price over the years. The focus of the proposed model is

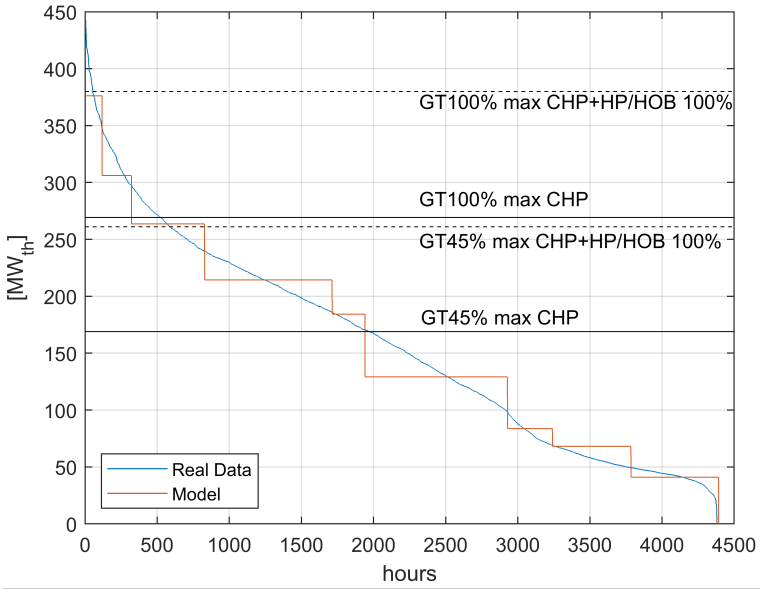
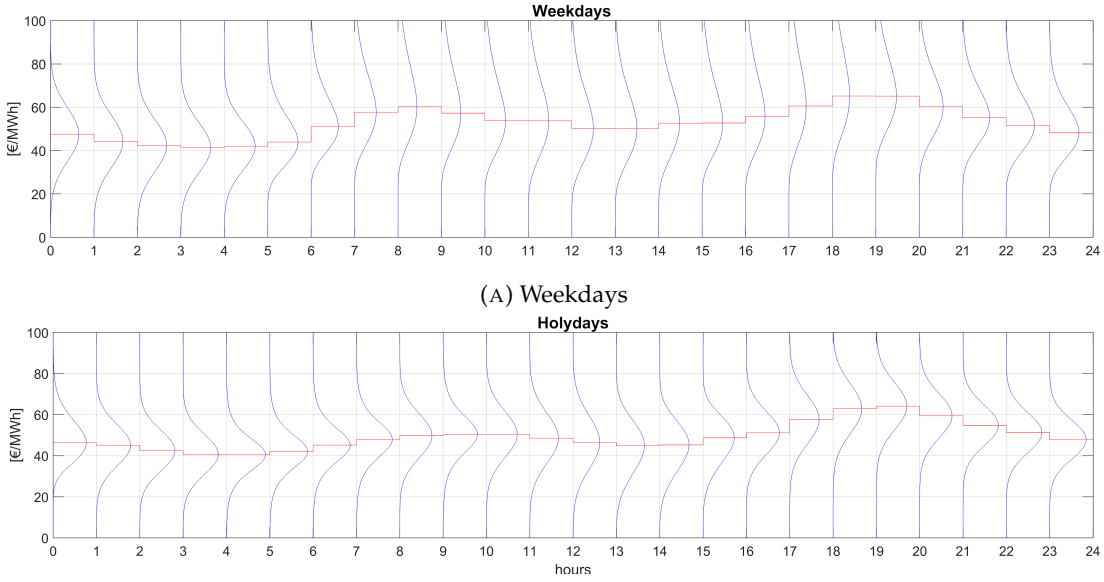


FIGURE 6.7: Heat demand profile, real and modelled

the long-term modeling of electricity price distributions approximating the hourly distribution during the two winter seasons of the years between 2016 and 2018 [198] with a specific probability density function.



(A) Weekdays

(B) Holydays

FIGURE 6.8: Electricity price probability density functions

In this study, the Single National Price (PUN) of electricity in Italy was considered, as this country was chosen as representative for the case study. To create the model the following steps were taken:

- the PUN was divided according to the "day-type": between weekdays and holidays, as it usually presents a different behavior in relation to that,

TABLE 6.4: Hourly PDF of the single national price (1h-12h)

Hour	Weekdays PDF	Main PDF parameters	Holidays PDF	Main PDF parameters
1	Normal	μ = 47.43 σ = 12.26	Gamma	θ = 20.83 k = 2.34
2	Normal	μ = 44.22 σ = 12.00	Normal	μ = 45.09 σ = 10.18
3	Normal	μ = 42.33 σ = 11.86	Normal	μ = 42.54 σ = 9.85
4	Normal	μ = 41.32 σ = 11.63	Normal	μ = 40.63 σ = 9.87
5	Normal	μ = 42.06 σ = 11.22	Logistic	μ = 40.54 δ = 5.55
6	Gamma	θ = 15.93 k = 2.95	Logistic	μ = 42.03 δ = 5.71
7	Inverse Gaussian	μ = 55.83 λ =931.92	Logistic	μ = 45.16 δ = 5.85
8	Inverse Gaussian	μ = 64.38 λ =878.16	Logistic	μ = 47.92 δ = 6.04
9	Inverse Gaussian	μ = 68.62 λ =787.85	Normal	μ = 49.71 σ = 10.75
10	Inverse Gaussian	μ = 65.77 λ = 714.97	Logistic	μ = 50.40 δ = 6.32
11	Inverse Gaussian	μ = 61.77 λ = 676.06	Normal	μ = 50.34 σ = 11.25
12	Gamma	θ = 11.16 k = 5.29	Normal	μ = 48.55 σ = 12.09

- the weekday and holiday PUNs were divided hour by hour, and the data corresponding to the same hour and day-type were then gathered,
- the gathered data were then analyzed and a MATLAB function was created to evaluate the Probability Density Function (PDF) which would best fit the data, according to the log likelihood criteria [199].

In Table 6.4 and 6.5 the PDF obtained and their characteristic parameters are listed, divided hour by hour. In Figure 6.8 the visual representation of the PDFs along the day, both for weekdays and holiday, are shown and the trend of the PDFs' peak is highlighted (red line). The symbols used to represent the PDF parameters are listed in the following with their corresponding PDF [18]:

- **Normal**: mean μ and standard deviation σ
- **Gamma**: shape parameter k and scale parameter θ
- **Inverse Gaussian**: mean μ and shape parameter λ

TABLE 6.5: Hourly PDF of the single national price (13h-24h)

Hour	Weekdays PDF	Main PDF parameters	Holidays PDF	Main PDF parameters
13	Gamma	θ = 10.95 k = 5.05	Normal	μ = 46.45 σ = 12.98
14	Gamma	θ = 11.06 k = 4.99	Weibull	b= 4.08 a= 48.10
15	Gamma	θ = 11.28 k = 5.12	Weibull	b= 4.28 a= 48.13
16	Inverse Gaussian	μ = 60.73 λ = 644.20	Weibull	b= 4.88 a= 51.03
17	Inverse Gaussian	μ = 64.25 λ = 678.49	Normal	μ = 51.09 σ = 10.48
18	Inverse Gaussian	μ = 70.68 λ = 686.42	Normal	μ = 57.62 σ = 11.77
19	Inverse Gaussian	μ = 74.60 λ = 830.07	Normal	μ = 62.93 σ = 12.06
20	Inverse Gaussian	μ = 71.60 λ = 1130.85	Normal	μ = 64.09 σ = 11.37
21	Gamma	θ = 18.41 k = 3.47]	Normal	μ = 59.47 σ = 10.70
22	Gamma	θ = 19.55 k = 2.98	Normal	μ = 54.77 σ = 9.80
23	Gamma	θ = 18.86 k = 2.89]	Normal	μ = 51.21 σ = 9.75
24	Gamma	θ = 17.85 k = 2.86	Normal	μ = 47.81 σ = 9.49

- **Logistic**: mean μ and scale parameter δ
- **Weibull**: shape parameter b, scale parameter a

It is interesting to notice that for weekdays, the Normal distribution is the best fit for the first hours of the day, while it is highly representative of the PUN behavior during holidays for most of the hours, so this PDF appear to be representative of low price periods. As concerns weekdays, Inverse Gaussian is representative of the morning hours and again the late afternoon-evening hours, where the peak prices are usually located (as can also be observed in Figure 6.8A). The Gamma distribution seems to have the effect of leading the transition between those different periods of the day, covering the "middle price period". During holidays instead, the Logistic PDF is selected to represent the morning hours, meaning that the related prices present heavier tails than Normal distribution.
The PDF were then expressed as a function of their mean, so that they could then be used to represent a different PUN behavior, assuming however that the shape

expressed by the PDF and by the standard deviation would remain unchanged over the years. The selection of the best fitting PDF and their parameters estimation was made considering three winter seasons, through a cross-validation procedure. The average error in the test phase is of 4.85% (2.90 €/MWh) in weekday and 7.4% (3.78 €/MWh), making the model suitable for the proposed analysis.
As stated before, the plant was assumed to operate for a 20-year life span, between 2020 and 2040. Three different PUN scenarios were considered [200]: a constant trend, an increasing and a decreasing trend. While an increase of the PUN is expected due to the forecasted increase of the gas cost, the constant and the decreasing PUN price scenarios are representatives of an electricity market with a high share of RES production, leading to a decoupling of the electrical price variation with respect to the gas price. As for gas, since linear trends were adopted, the average values are equal to the average of the distribution in 2030 for the three scenarios.
In the average PUN scenario, the average of the distributions remains the same over the years as the mean corresponding the winter season 2018/2019 (i.e. 60.77 €/MWh for weekdays and 55.18 €/MWh for holidays). In the High PUN scenario an increase of the average price of electricity of 2.5%/year, compared to the current value, is forecasted, resulting in an average of 75.96 €/MWh for weekdays 68.97 €/MWh for the others. In the low PUN scenario, a decrease of average price of electricity of 2.5%/year, compared to the current value, is considered with an average of 45.58 €/MWh for weekdays and 41.38 €/MWh during weekends and holidays. These values were used as the mean to build the PDF listed in Table 6.4 and 6.5.

6.4.4 Evaluation of economic profitability

The approach of the real CHP power plant is oriented to satisfy the heat demand and generate the electrical power that allows one to maximize profits or minimize losses. So, the appropriate level of electrical generation depends exclusively on the economic parameters described in the previous subsection. The model assesses all the operating conditions to satisfy the given thermal demand and then selecting the most profitable operating condition. So, an iso-thermal line, parallel to the y-axis, is ideally drawn on the Iron diagrams (Figure 6.5), and the operating and performance parameters of the plant are evaluated for each gas turbine load linearly interpolating between the known points from the model. The optimization algorithm is performed also on the maximum steam extraction, the maximum heat pump, or HOB, power lines. In all these conditions the economical incomings are computed as follow:

$$\begin{aligned} I = P_{th} p_{heat} + P_{el} PUN - L\dot{H}V \left(C_{gas} + eC_{CO_2}\right) \\ -O\&M_{CCGT_{var}} P_{el,CCGT} - O\&M_{HP/HOB_{var}} P_{th,HP/HOB} \end{aligned} \tag{6.3}$$

where I is the incoming in the unit of time excluding the fixed $O\&M$ costs, P_{th} and P_{el} are the electrical end the thermal power output of the plant,$L\dot{V}H$ is the primary energy consumption, e is the emission factor, p_{heat} is the price of heat, as computed from eq. 6.2, and C_{CO_2} is the cost of CO_2, as described in the Section 6.4.2.
Selected the point maximizing I, then the annual cash flow is evaluated summing for the whole season and considering the fixed O& M costs:

$$CFN_{HP} = \sum_{i}^{n} I_i r_i - O\&M_{CCGT-HP_{fix}} \tag{6.4}$$

$$CFN_{HOB} = \sum_{i}^{n} I_i r_i - O\&M_{CCGT-HOB_{fix}} \tag{6.5}$$

Since there are three possible daily thermal load profile (Table 6.3) and two kinds of day (weekdays and holidays) are considered, the season is subdivided in 144 standardized hours characterized by a thermal load and an electricity price. For each of these the occurrence r is computed, defined as the number of hours during a season on which the condition occurs. Thus Equations 6.4 and 6.5 sum over the n = 144 standardized hours the hourly incoming I, Equation 6.3, times the occurrence r in order to model the whole cold season period, 4368 h. The fixed and variable Operating and Maintenance costs for the CCGT, the HP and the HOB, figure in the equations 6.3, 6.4 and 6.5 according to the subscripts reported in the nomenclature.

6.5 Results

This Section reports the results of the analysis performed, according to the methodology described in Section 6.2. The number of samples for which the Monte Carlo simulation gave an acceptable MSPE value turned out to be 15000. Firstly, the average HOB and HPs' load distribution is presented in order to validate the off-design assumptions and guarantee the results reliability. Then, in the first subsection the results related to the economic viability and profitability of the proposed solutions are reported, while the second compares the CCGT-HP and CCGT-HOB under the carbon emission reduction perspective. The last one presents a sensitivity analysis with respect to the carbon cost.
Figure 6.9 shows how the HPs and the HOB mean value loads are distributed in the Average PUN scenario. It is possible to appreciate how the average load of each single HP is never lower than 60%, and the same minimum value is observed in the two other PUN scenarios, thus the assumption related to the off-design operations made in Section 6.3.1 results to be validated. Also the minimum HOB load is never lower than 10% in any scenario, justifying the application of the Satyavada et al. model [190].
Table 6.6 integrates the information of Figure 6.9, providing a comprehensive report of the yearly operating hours for the HPs and HOB in every considered PUN scenario. In high PUN scenario the CCGT works at higher load, in most of the cases without requiring any integration in order to satisfy the thermal demand. To quantify this data, it is important to remark that the seasonal overall number of hours is 4368. Decreasing the price of electricity, it results to be preferable, especially for the CCGT-HP solution, to lower the GT load integrating the HPs or the HOB to provide the required thermal power. This is directly reflected on an increase of yearly operating hours of these components.

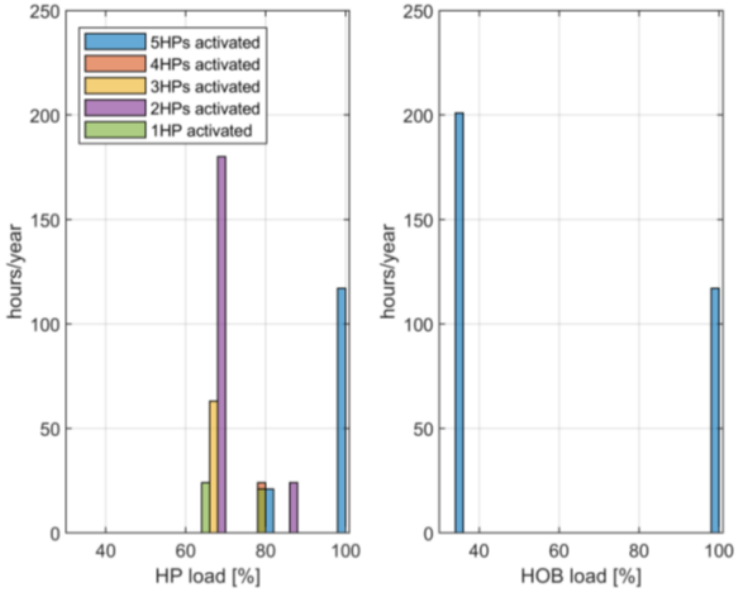


FIGURE 6.9: HPs (left) and HOB (right) yearly off-design operating hours (average PUN scenario)

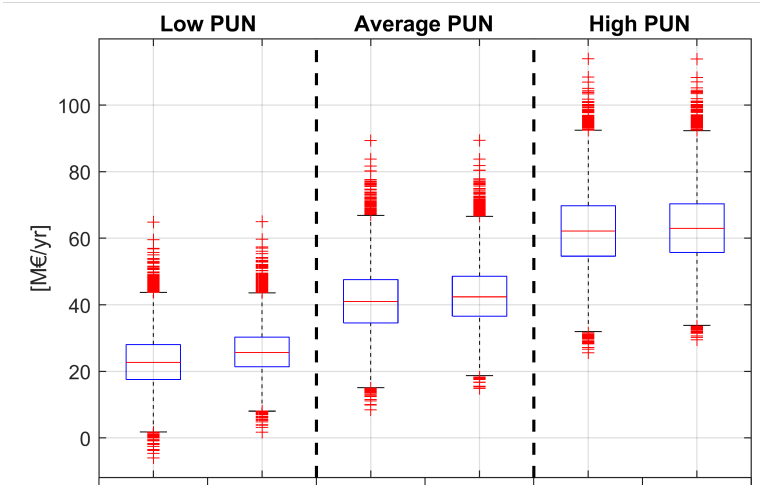
TABLE 6.6: HPs and HOB operating hours

	HP	HOB
Low PUN	1587	486
Average PUN	474	318
High PUN	318	318

6.5.1 Economics

Looking at the three scenarios on which the analysis was carried out, it is important to observe how the annual cash flow (as shown by Figure 6.10) always has a median value, red line, higher for the CCGT-HP than the CCGT-HOB. It is also important to point out that the integrated heat pump system seems to be less affected by market uncertainties, as shown by the distance betwen the 1st and the 3rd quartilie, blue box, and the 95th and 5th percentile (exact values are reported in table immediately below the figure). This is strictly connected to the fact that, in order to achieve higher thermal power levels, electrical power is employed instead of primary energy and so, in case of low electricity prices, the gas consumption can be reduced by consuming a cheaper source. Thus, the CCGT-HP solution is less sensitive to the market uncertainties considered. This trend becomes more evident when the price of electricity decreases, which would represent the most critical condition from the sustainability point of view, due to lower annual cash flow. In fact, while in the High PUN scenario the expected cash flow value for the CCGT-HP is 1.3% higher than CCGT-HOB, this advantage increases up to 13.2% in the low PUN scenario. This shows the potentialities of the CCGT-HP in critical scenarios: while the variability (expressed as the difference between the 75th and 25th percentiles) in the High PUN scenario is 3.18% lower than the CCGT-HOB solution, for the low PUN scenario the cash flow variability reduction becomes equal to 15.25%.

Nevertheless, as reported in the Section 6.4.1, the capital expenditures are consid-



	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP
Median	22.6653	25.6654	40.954	42.3909	62.1438	62.9575
75^{th}-25^{th}	10.4861	8.8865	13.0059	12.018	15.133	14.6513
95^{th}-5^{th}	25.7499	21.9028	31.5634	29.1072	36.4951	35.2244

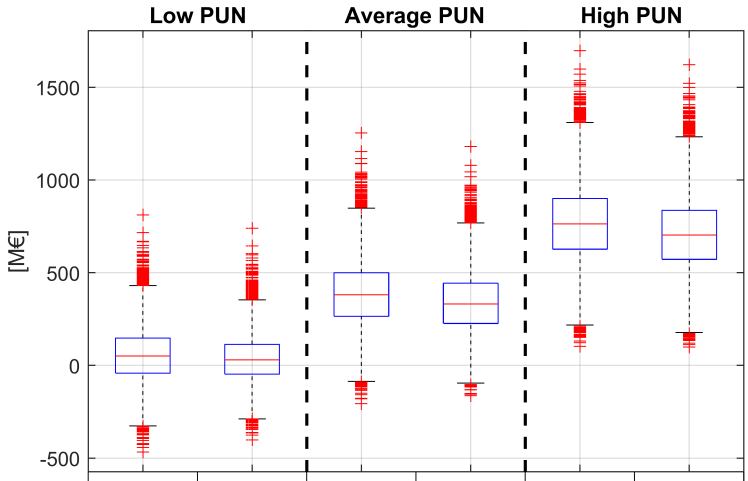
FIGURE 6.10: Annual cash flow distribution

erably higher for the solution integrating the heat pump, consequently, despite the higher annual cash flow, the CCGT-HP presents a lower expected Net Present Value (NPV). In particular, the NPV presented in Figure 6.11 is calculated as:

$$NPV = -\text{CAPEX} + \sum_{j=1}^{20} \frac{CFN_j}{(1+d)^j} \tag{6.6}$$

Where d is the discount rate set to 1% over a 20 years plant's lifetime. The Net Present Value shows a dependence on the variability of both price of electricity and cost of gas, like the annual cash flow. From Figure 6.11 it can be observed that the High PUN scenario would be particularly beneficial in terms of NPV for both solutions (CCGT-HP and CCGT-HOB), presenting an expected NPV higher than 700 M€thanks to the higher revenues of the electrical production. This high electricity price gives benefit particularly to the CCGT-HOB system, in which all electrical production can be sold at market price. On the other hand, in a low PUN scenario, the CCGT-HP presents just a slightly lower average NPV, but with a lower variability. Nevertheless, in such a critical situation, both systems present negative values for the first NPV quartile, with a risk of not even covering the investment costs (i.e. 358.64 M€for CCGT-HOB and 433.12 M€for CCGT-HP). This risk is quantified in 35.9% of probability for the CCGT-HOB and in 39.8% for CCGT-HP.
The economic conditions foreseen in this analysis indicate that the HOB solution tends to be more profitable in high electrical price scenarios. On the other hand, in the event of decreasing electrical prices, as in case of high share of renewables, the economic performance would be comparable, but with a lower variability associated to the CCGT-HP solution. The NPV increases in case of higher cash flows: this can be obtained in sites where a larger annual thermal demand exists (e.g. in countries colder than northern Italy) or by increasing the heat price paid by the consumers. Moreover, it is also important to remark that with the advent of the most recent generations of DHN, and the following decrease in temperature levels, the heat pumps

will be able to work with a higher COP with a benefit to the thermodynamic efficiency and the economic profitability.



	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP
Median	50.3678	30.0315	380.3975	331.8517	762.7793	702.9882
75^{th}-25^{th}	189.2279	160.3617	234.6992	216.8721	273.0838	264.3901
95^{th}-5^{th}	464.672	395.2474	569.5788	525.2551	658.574	635.6446

FIGURE 6.11: New present value distribution

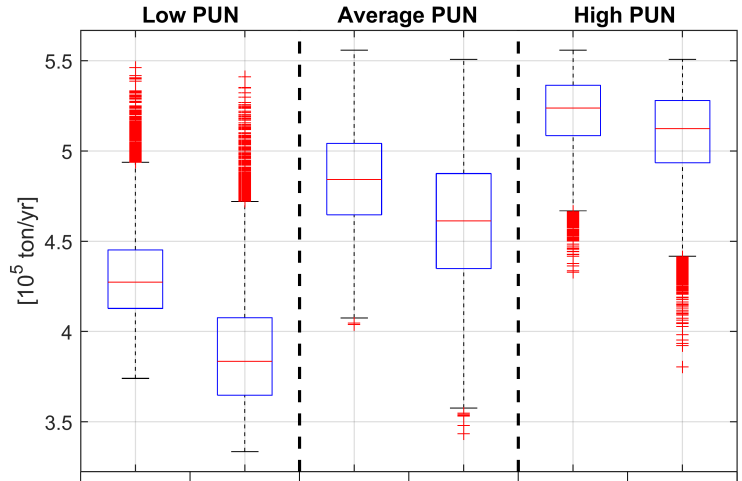
6.5.2 Environmental Impact

Under the GHG emission perspective, Figure 6.12 shows that the CCGT-HP presents a reduced amount of emitted carbon dioxide. The reason is twofold: on one hand the CCGT-HP system works with higher global efficiency, as highlighted by Figure 6.13, on the other hand the electricity production is also lower since lower electricity prices lead the system to run closer to the minimum load that enables the fulfillment of the thermal demand.

Increasing the price of electricity, the emitted carbon dioxide distribution (Figure 6.12) moves towards higher values (increments of $0.96 \cdot 10^5$ and $1.29 \cdot 10^5$ ton/year for CCGT-HOB and CCGT-HP respectively) and shows a lower standard deviation for both the solutions proposed. This is due to the fact that the power plant increases the number of hours at which it operates in full load conditions, where it can provide a higher thermal power without any integration, consequently increasing the number of hours in which there is no difference between the CCGT-HOB and the CCGT-HP environmental behavior.

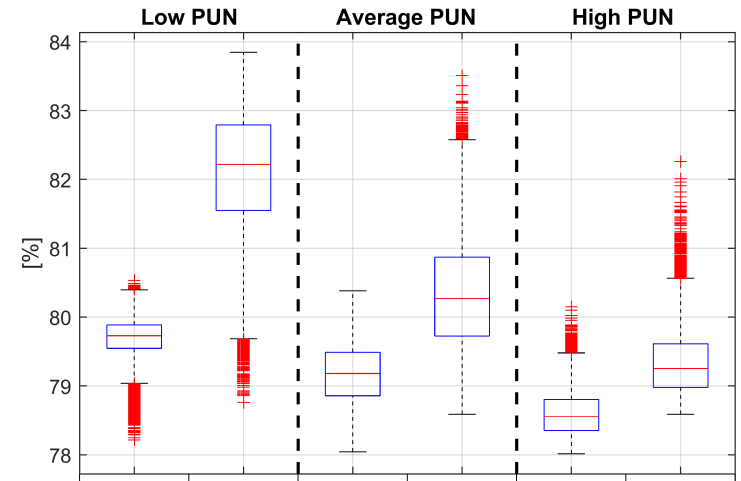
It is also interesting to observe how the carbon dioxide emission distributions, Figure 6.12, as well as the average global efficiency, Figure 6.13, in the Low and High PUN scenario present opposite skewness. This is due to the electricity price which leads the power plant always to operate at the minimum or at full load. Thus, decreasing or increasing further the PUN does not affect the power plant's emissions.

The larger variabilities registered in yearly average global efficiency and CO_2 emissions for CCGT-HP are due to the larger enhancement in global efficiency that can be reached operating the HP with respect to the HOB. This reflects also on the CO_2 emissions.



	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP
Median	4.2738	3.8353	4.8417	4.6123	5.2375	5.1241
75^{th}-25^{th}	0.32366	0.42942	0.3957	0.52642	0.27881	0.3453
95^{th}-5^{th}	0.81593	1.0609	0.89933	1.1965	0.64325	0.8031

FIGURE 6.12: Carbon dioxide emission distribution



	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP
Median	79.7291	82.2169	79.1804	80.2693	78.5542	79.2526
75^{th}-25^{th}	0.34002	1.2427	0.63052	1.1444	0.45182	0.63449
95^{th}-5^{th}	0.90286	2.9215	1.3898	2.6345	1.0535	1.527

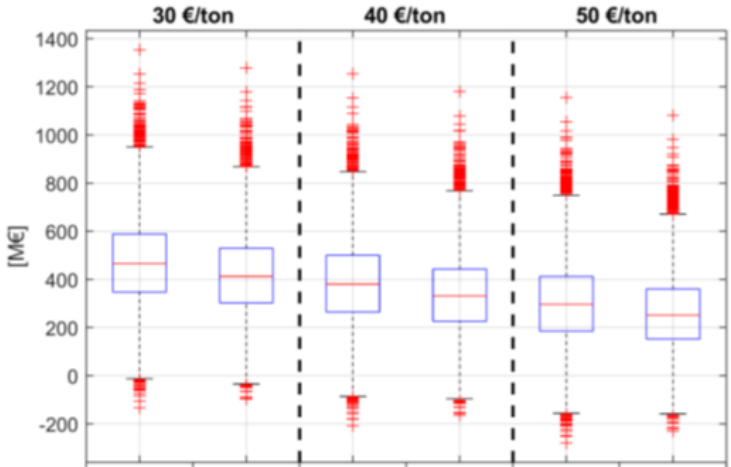
FIGURE 6.13: Average global efficiency distribution

6.5.3 Carbon cost sensitivity

Finally, a sensitivity analysis was performed with respect to the emitted carbon dioxide cost. The effect of the cost of emitted CO_2 is similar to the cost of fuel in calculating the overall cost of primary energy consumption: an increment of 10 €/ton of carbon dioxide correspond to an increment of 1.83 €/MWh for the fuel cost. The main difference is that CO2 cost change, in the proposed model, is not reflected on the thermal customers but is repaid by the CHP annual income affecting the profit margin. Table 6.7 presents the yearly emission value: from the policy/environmental point of view an increase of the cost of the emitted CO_2 is appreciable, since it leads

to a decrease of the environmental impact, keeping fixed the satisfied heat demand. In fact, a higher carbon cost leads the plant to operate at partial load saving primary energy. Nevertheless, as shown by Figure 6.14, this implies an increase of the risk of not paying the investment cost back, and it is discouraging for a real plant owner to keep the plant operating in such market conditions.

In fact, an increase in CO_2 cost is reflected on a decrement in final profitability without significative differences between the CCGT-HP and CCGT-HOB. Averagely for an increase of 10 €/ton a 18.2% decrement in NPV occurs, while for the CCGT-HOB this decrement is of 19.4%.



	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP	CCGT-HOB	CCGT-HP
Median	465.531	412.7721	380.3975	331.8517	296.3682	252.2474
75^{th}-25^{th}	241.3657	225.9687	234.6992	216.8721	227.5821	208.1459
95^{th}-5^{th}	587.0699	549.2702	569.5788	525.2551	556.1311	507.2936

FIGURE 6.14: NPV distribution for different carbon cost scenarios (average PUN)

TABLE 6.7: Ton of CO_2 emitted for different carbon cost senarios (average PUN)

	30 €/ton		**40 €/ton**		**50 €/ton**	
	CCGT HOB	**CCGT HP**	**CCGT HOB**	**CCGT HP**	**CCGT HOB**	**CCGT HP**
Median	$4.94 \cdot 10^5$	$4.75 \cdot 10^5$	$4.84 \cdot 10^5$	$4.61 \cdot 10^5$	$4.74 \cdot 10^5$	$4.47 \cdot 10^5$
$75^{th} - 25^{th}$	$0.402 \cdot 10^5$	$0.525 \cdot 10^5$	$0.396 \cdot 10^5$	$0.526 \cdot 10^5$	$0.385 \cdot 10^5$	$0.523 \cdot 10^5$

6.6 Concluding remarks

In this work it was presented a methodology to include the uncertainties of market forecast within the long-term thermo-economic assessment of Energy Systems. A

test case was developed on a new solution for CHP generation, integrating a Combined Cycle Gas Turbine, CCGT, with a flue gas condensing heat pump, and comparing it with a conventional coupling of the same CCGT with a Heat-Only Boiler. The performance of the integrated systems, based on a 400 MW F Class CHP CCGT Gate Cycle with detailed calculations, were evaluated through a model developed in MATLAB. Both systems enhanced the heat production by 110MWth. While the integration with the heat pump, exploiting also the latent heat from the flue gases, allows one to increase the global efficiency, estimated on LHV, from 89.5% to 101.2%, the Heat-Only Boiler integrated system reaches 89%, since additional fuel is required.
To properly evaluate its feasibility from the economic point of view a thermo-economic analysis under uncertainty of the system was performed, modeling the cost of gas and single national price of electricity with a series of probability density functions to recreate the daily profile. A similar approach is followed for the gas cost, where the uncertainties related to the forecast capability are considered by creating a normal distribution with a proper standard deviation.
The Italian energy market was selected as reference for the case study, however the methodology employed could easily be scaled to different markets and systems. In particular, the electricity market model described in Section 6.4.3 would only require providing a valuable set of data on which the PDF could be selected. When considering instead the creation of the model for the thermal demand, explained in Section 6.4.2, this would obviously require a proper understanding of the market analyzed.
A Monte-Carlo simulation was performed on the two systems, and the outputs were then analyzed, leading to the following main conclusions:

- despite a higher annual cash flow for the CCGT-HP solution, the higher investment costs of this solution ended up in a lower NPV when compared to the CCGT-HOB solution. However, a lower variability of the CCGT-HP solution was also highlighted, ending in a more robust solution, especially in the more critical situation, the one with decreasing electrical prices.

- From the environmental point of view the CO2 emission was analyzed. As concerns this variable, the CCGT-HP presents lower emission than the CCGT-HOB, because the thermal capacity of the CHP power plant is extended exploiting the flue gas latent heat by the heat pump, thus increasing global efficiency, rather than burning additional fuel as in the HOB.

The proposed integration of HP and CCGT appears to be interesting under a generation mix with a strong presence of renewable generators which brings to an electrical price uncorrelated to the gas cost (Low PUN scenario): in this case the revenues related to electricity market are no more covering the production cost and the system profitability is just related to the thermal demand and the depreciated electricity can be used to generate heat through the HP. Under this condition, the CCGT-HP could be of interest for municipalities interested in fulfilling the local thermal demand guarantying lower emissions and reducing the risk over the years of deviating from expected plan.
Finally, it interesting to highlight that economic performances and environmental-thermodynamic performances have an inverse correlation with the electrical prices: the NPV grows with the electrical market price for both systems while efficiency reduces, and the CO_2 emissions increase due to the higher use of the CCGT at full load. The cost of emitted CO_2 has a similar impact: an increase of this parameter, if

repaid internally by the CHP, will bring to a decrease of carbon emission but also to a reduction of system profitability.
Generalizing the results: the NPV could increase, for both systems, in case of higher cash flows: this can be obtained in sites where a larger annual thermal demand exists (e.g. in countries colder than northern Italy) or by increasing the heat price paid by the consumers. Moreover, it is also important to remark that with the advent of the most recent generations of DHN, and the following decrease in temperature levels, the heat pumps will be able to work with a higher COP with a benefit to the thermodynamic efficiency and the economic profitability. Finally, since a lower average electricity price is beneficial to the proposed CCGT-HP, market with excess of production from renewable energy system could be interested to such an application.

References

[18] R. Ghanem, H. Owhadi, and D. Higdon. *Handbook of uncertainty quantification*. Ed. by Springer. Berlin, Germany, 2017, pp. 1–2053. ISBN: 9783319123851. DOI: 10.1007/978-3-319-12385-1.

[77] L. Cassettari, R. Mosca, and R. Revetria. "Experimental error measurement in monte carlo simulation". In: *Handbook of Research on Discrete Event Simulation Environments: Technologies and Applications*. 2009, pp. 92–142. ISBN: 9781605667744. DOI: 10.4018/978-1-60566-774-4.ch006.

[165] European Commission. *A policy framework for climate and energy in the period from 2020 to 2030*. Tech. rep. 2014.

[166] Eurostat. *Energy, transport and environment indicators. 2018 edition*. 2018, p. 232. ISBN: 978-92-79-96509-8. DOI: 10.2785/94549. URL: htttp://dx.doi.org/10.2785/94549.

[167] M. Minutillo, A. Perna, and A. Sorce. "Combined hydrogen, heat and electricity generation via biogas reforming: Energy and economic assessments". In: *International Journal of Hydrogen Energy* 44.43 (2019), pp. 23880–23898. ISSN: 03603199. DOI: 10.1016/j.ijhydene.2019.07.105.

[168] D. Bellotti, A. Sorce, M. Rivarolo, and L. Magistri. "Techno-economic analysis for the integration of a power to fuel system with a CCS coal power plant". In: *Journal of CO2 Utilization* 33 (2019), pp. 262–272. ISSN: 22129820. DOI: 10.1016/j.jcou.2019.05.019.

[169] D. Bellotti, L. Cassettari, M. Mosca, and L. Magistri. "RSM approach for stochastic sensitivity analysis of the economic sustainability of a methanol production plant using renewable energy sources". In: *Journal of Cleaner Production* 240 (2019). ISSN: 09596526. DOI: 10.1016/j.jclepro.2019.117947.

[170] A. D. Smith, P. J. Mago, and N. Fumo. "Benefits of thermal energy storage option combined with CHP system for different commercial building types". In: *Sustainable Energy Technologies and Assessments* 1.1 (2013), pp. 3–12. ISSN: 22131388. DOI: 10.1016/j.seta.2012.11.001.

[171] B. Caldecott and J. McDaniels. "Stranded generation assets : Implications for European capacity mechanisms , energy markets and climate policy". In: *Stranded Assets Programme, SSEE, University of Oxford* January (2014), pp. 1–62.

[172] S. Cornot-Gandolphe. *Gas and Coal Competition in the EU Power Sector*. 2014.

[173] R. Wiltshire. *Advanced District Heating and Cooling (DHC) Systems*. 2015, pp. 1–364. ISBN: 9781782423959. DOI: 10.1016/c2014-0-01422-0.

[174] A. David, B. V. Mathiesen, H. Averfalk, S. Werner, and H. Lund. "Heat Roadmap Europe: Large-scale electric heat pumps in district heating systems". In: *Energies* 10.4 (2017). ISSN: 19961073. DOI: 10.3390/en10040578.

[175] R. Lund and U. Persson. "Mapping of potential heat sources for heat pumps for district heating in Denmark". In: *Energy* 110 (2016), pp. 129–138. ISSN: 03605442. DOI: 10.1016/j.energy.2015.12.127.

[176] F. Levihn. "CHP and heat pumps to balance renewable power production: Lessons from the district heating network in Stockholm". In: *Energy* 137 (2017), pp. 670–678. ISSN: 03605442. DOI: 10.1016/j.energy.2017.01.118.

[177] M. B. Blarke and E. Dotzauer. "Intermittency-friendly and high-efficiency cogeneration: Operational optimisation of cogeneration with compression heat pump, flue gas heat recovery, and intermediate cold storage". In: *Energy* 36.12 (2011), pp. 6867–6878. ISSN: 03605442. DOI: 10.1016/j.energy.2011.10.008.

[178] T. Ommen, W. B. Markussen, and B. Elmegaard. "Heat pumps in combined heat and power systems". In: *Energy* 76 (2014), pp. 989–1000. ISSN: 03605442. DOI: 10.1016/j.energy.2014.09.016.

[179] Q. Chen, K. Finney, H. Li, X. Zhang, J. Zhou, V. Sharifi, and J. Swithenbank. "Condensing boiler applications in the process industry". In: *Applied Energy* 89.1 (2012), pp. 30–36. ISSN: 03062619. DOI: 10.1016/j.apenergy.2010.11.020.

[180] B. Hebenstreit, R. Schnetzinger, R. Ohnmacht, E. Höftberger, J. Lundgren, W. Haslinger, and A. Toffolo. "Techno-economic study of a heat pump enhanced flue gas heat recovery for biomass boilers". In: *Biomass and Bioenergy* 71 (2014), pp. 12–22. ISSN: 09619534. DOI: 10.1016/j.biombioe.2014.01.048.

[181] M. Qu, O. Abdelaziz, and H. Yin. "New configurations of a heat recovery absorption heat pump integrated with a natural gas boiler for boiler efficiency improvement". In: *Energy Conversion and Management* 87 (2014), pp. 175–184. ISSN: 01968904. DOI: 10.1016/j.enconman.2014.06.083.

[182] Danish Energy Agency. *Technology Data for Energy Plants for Electricity and District heating generation*. Tech. rep. June. 2019, pp. 1–374.

[183] A. Molyneaux, G. Leyland, and D. Favrat. "Environomic multi-objective optimisation of a district heating network considering centralized and decentralized heat pumps". In: *Energy* 35.2 (2010), pp. 751–758. ISSN: 03605442. DOI: 10.1016/j.energy.2009.09.028.

[184] *Pump-Heat*. URL: www.pumpheat.eu (visited on 01/27/2021).

[185] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, and B. V. Mathiesen. *4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems.* 2014. DOI: 10.1016/j.energy.2014.02.089.

[186] A. Sorce, A. Giugno, D. Marino, S. Piola, and R. Guedez. "Analysis of a combined cycle exploiting inlet conditioning technologies for power modulation". In: *Proceedings of the ASME Turbo Expo*. Vol. 3. 2019, pp. 1–14. ISBN: 9780791858608. DOI: 10.1115/GT2019-91541.

[187] Friotherm. *Uniturbo 50FY – Centrifugal Compressor for Large Scale Refrigeration Plants and Heat Pumps*. Tech. rep.

[188] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort. "Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library coolprop". In: *Industrial and Engineering Chemistry Research* 53.6 (2014), pp. 2498–2508. ISSN: 08885885. DOI: 10.1021/ie4033999.

[189] A. society of heating refrigerating and air conditioning engineers. *1997 ASHRAE Handbook: Fundamentals*. 1997, p. 851.

[190] H. Satyavada and S. Baldi. "A novel modelling approach for condensing boilers based on hybrid dynamical systems". In: *Machines* 4.2 (2016). ISSN: 20751702. DOI: 10.3390/machines4020010.

[191] M. S. Peters, K. D. Timmerhaus, and R. E. West. *Plant Design and Economics for Chemical Engineers 5th Edition*. 2003. ISBN: 9789896540821. arXiv: arXiv:1011.1669v3.

[192] P. Capros et al. *EU Reference Scenario 2016 - A Policy Framework for Climate and Energy in the Period from 2020 to 2030*. Tech. rep. 2016, p. 27. DOI: 10.2833/9127.

[193] Gestore dei Servizi Energetici. *Assessment of the National and Regional Potential for the Application of High-Efficiency Cogeneration and Efficient District Heating*. Tech. rep. 2016.

[194] Autorità Garante della Concorrenza e del Mercato. *Indagine Conoscitiva Sul Settore Del Teleriscaldamento*. Tech. rep. 2014.

[195] H. Li, J. Song, Q. Sun, F. Wallin, and Q. Zhang. "A dynamic price model based on levelized cost for district heating". In: *Energy, Ecology and Environment* 4.1 (2019), pp. 15–25. ISSN: 23638338. DOI: 10.1007/s40974-019-00109-6.

[196] Gestore Mercati Energetici. *Dati Storici Mercati*. Tech. rep. 2018.

[197] M. dell'ambiente e della tutela del territorio e del mare. *Decreto Legislativo n.30 Del 2013, Tabella Allegata Dei Parametri Standard Nazionali per Il Monitoraggio e La Comunicazione Dei Gas Ad Effetto Serra*. 2019.

[198] TERNA. *Dati statistici sull'energia elettrica in Italia*. Tech. rep. 2018. URL: http://www.terna.it/SistemaElettrico/StatisticheePrevisioni/DatiStatistici.aspx.

[199] R. A. Fisher. "On the mathematical foundations of theoretical statistics". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222.594-604 (1922), pp. 309–368. ISSN: 0264-3952. DOI: 10.1098/rsta.1922.0009.

[200] European Commission. *EU Reference Scenario 2016 - Energy, transport and GHG emissions - Trends to 2050*. Tech. rep. 2016. URL: https://ec.europa.eu/energy/sites/ener/files/documents/20160713draft{_}publication{_}REF2016{_}v13.pdf.

Chapter 7

Conclusion and future research directions

This chapter provides a summary of the research presented in the previous chapters and the significant results obtained. It further includes suggested directions for future work that could be pursued in this area of research

7.1 Conclusion

Mechanical systems operate intrinsically under uncertainty, principally related to the operation and design condition. On one hand, many problems, which complexity posed serious limitations in the past, are now entering a domain of computational time for which the stochastic simulation and the UQ analysis become possible, thanks to recent developments in the field of high-performance computing. On the other hand, new numerical methods for UQ are appearing, which provide better accuracy while requiring a lower computational burden. These two trends suggest that UQ will become increasingly important and applied in the near future.
The presented framework for UQ represents a valuable tool in the analysis of advanced energy systems subject to uncertainties. These analyses have been traditionally performed using probabilistic methods, some of which have been presented in this work. The computational bottleneck in UQ is often located in the computation of the forward model, which is required for every step of the UQ framework but the quantification of the input uncertainties – unless these are being characterized using probabilistic inversion. This forward model describes the physics of the problem, and despite big progresses in the field of high-performance computing, the solution of many of these models is computationally challenging. This fact limits the applicability of traditional probabilistic methods, which often require hundreds or thousands of evaluations of the forward model. This justifies the application of response surface methodologies or metamodels, which try to find an accurate approximation of the forward model to perform the analysis of interest. In this work, two techniques for uncertainty quantifications have been investigated: the Monte Carlo simulation and the Response Sentivitiy Analysis. The first method in particular requires a huge computational effort, however his strenghts in terms of uncertainty quantification are undoutbtful and for this reason the author has decided to make use of the response surface methodology to approximate the models of advanced

energy systems.
Hence, the aim of this work was to provide a valuable method for a comprehensive analysis under uncertainty of advanced energy systems, with a particular focus on hybrid systems, considering the impact that such uncertainties would have on the system performance and revenues. The focus was laid out on aleatory uncertainties caused by natural, unpredictable variations of parameter values, initial and boundary conditions, machine maps and of the geometry itself. Due to the stochastic nature of the considered uncertainties, a stochastic approach modeling the behaviour of the uncertain quantities was chosen. The additional information of the uncertain parameters given by the stochastic model can be used to reduce the computational effort arising by the computation of statistics of the quantities of interest. The numerical results demonstrate the potential of exploiting the stochastic information of the input uncertainties. In the next points, the main steps performed during this thesis for the creation of a valuable method for such analyses are briefly summarized:

1. Starting from a "simple" recuperator model for micro gas-turbine, the application of a sampling method as Monte Carlo and an approximate method such as Response Sensitivity Analysis was carried out. These methods have allowed to get insights on the impact of uncertainties related to design parameters on the cost and volume, which are critical aspects of this kind of component.

2. The application of a sampling method to more complex energy system was not feasible due to a too long computational time required. For this reason the main response surface methodology techniques have been studied, analyzed and finally applied to a solid oxide fuel cell hybrid system in order to create a metamodel which could resemble the behaviour of the system. These response have allowed to apply an expensive method such as Monte Carlo on the system and performing a multi-objective optimization, obtaining valuable insights in terms of performance and economic indicators. The range of produced power and of payback period is extremely important when commissioning this system as it can give the owner an idea of the revenues of the system and of the operating conditions to expect.

3. A first step in the analysis under uncertainties of complex energy systems such as combined cycle was made, creating a model to extimate the behaviour of future market conditions. Analyzing the impact of market uncertainties in terms of electricity and gas prices on the revenues of such system and on its emissions.

The creation of a response surface assumes primary importance when performing an uncertainty analysis or optimization on more complex systems. Even if the choice of the number of parameters and of the proper response surface methodology technique can be challenging, it is a step that needs to be taken.
From a quantitative point of view, the main results obtained in this dissertation are outlined below:

- The **recuperator analysis** showed the potential of coupling a detailed sizing tool with uncertainty quantification methods. The uncertainties on recuperator key parameters, such as pressure drop and effectiveness was considered.Monte Carlo simulation has allowed the author to evaluate the probability density

function of the monitored outputs and has highlighted a high variability of cost and volume. The application of the RSA method has allowed to further investigate the root cause, highlighting the effectiveness as the main parameter affecting the results, while the hot side pressure drop presented a small influence and the cold side pressure drop no relevant impact at all.

- The **analysis of hybrid systems** under uncertainties was then presented, considering a case study of a turbocharger hybrid system featuring a Solid Oxide Fuel Cell (SOFC) stack. In all cases it was chosen to make use of a response surface metamodel (Central Composite Face Centered (CCF) with Least Square Method (LSM)) in order to simplify the Matlab-Simulink model from a computational point of view to be able to perform an uncertainty quantification study which is normally computationally expensive. The work was divided into three main parts:

  1. **Analysis of the propagation of the uncertainties on the performance of the system**. Seven different inputs have been considered affected by uncertainty and their effects on power produced, Steam-To-Carbon Ratio (STCR), percentage of mass flow through the wastegate valve and turbocharger rotational speed has been evaluated. Results showed that the fuel cell ohmic resistance, the CO_2 percentage in fuel and the ejector diameter are the factors which present the strongest impact on performance, thus they are crucial to be determined accurately, with the lowest standard deviation possible, to guarantee the power and the efficiency desired. The application of Monte Carlo simulation to the different response surfaces has allowed to estimate the mean, the standard deviation and the coefficient of variance of the monitored outputs. In particular, by comparing the coefficient of variance (COV) of the outputs to the input ones, it has been possible to highlight that the model, hence the plant, is able to mitigate the propagation of uncertainties, as the COV of the outputs is lower than the COV of the inputs.
  2. **Design under uncertainty of the system and robust design**. To evaluate how would the sistem perform with the replacement of the SOFC stack and of the recuperator, the area of the solid oxide fuel cell (A_{cell}), the current density (J_{cell}) and the recuperator surface (A_{REC}) have been considered as factors to build the response surfaces. The effect of uncertainties in the turbocharger maps in terms of turbine and compressor efficiency (η_t, η_c) and in the ohmic losses of the SOFC stack (k_{ohm}) were also considered when building the response surfaces. The analysis was carried out to evaluate the impact of the uncertainties on the performance and on the economic profitability of the system, considering three differet market scenarios: Italy, United Kingdom and New York.
     The response surfaces highlighted that the net efficiency and net power depended mostly on the cell area and stack current density, as the power produced is related to the SOFC stack performance only. In particular, the highest net power production would be achieved using a larger stack (high A_{cell} and high J_{cell}), while the net efficiency presented a maximum for low J_{cell}, as this would impact directly the ohmic losses and then the efficiency. From a first-order economic analysis it was highlighted that United Kingdom presented the lowest payback period and internal rate of return, followed by New York and then by Italy. This is related to the

fact that Italy presents a very high cost of natural gas, which is almost double the one of United Kingdom and New York.
The polynomial response surfaces of net power and capital costs were then used to perform a multi-objective optimization with genetic algorithm and a Pareto front was obtained. The Pareto front allowed to evaluate the optimum solutions in terms of costs and power production, pointing out the two "top power production" solutions. A Monte Carlo simulation was performed on such response surfaces to evaluate the coefficient of variance and so the robustness of the solutions. This analysis showed that the point with the highest net power production presented a lower coefficient of variance both when considering power and both when considering efficiency, resulting in a more robust solution. This result is particularly interesting from the point of view of the manufacturer, as it would mean that choosing this solution, despite a slightly higher capital cost, would guarantee less uncertainties towards customers, a more reliable design production less affected by uncertainties and tolerances of the system.

3. **Off-design analysis of the system subject to technological uncertainties**. The solution with the highest internal rate of return considering the Italian market scenario from the previous work was chosen as the design point. An off-design analysis of such design point was then performed considering three different loads: 100%, 85% and 70% electrical power output. The response surfaces of net power and net efficiency were then built for each load percentage, using the uncertainties as factors to evaluate their impact on system performance. From this analysis it was highlighted that one of the points of 100% load condition, estimated by the metamodel, was not feasible, as the uncertainties were contributing to raise the SOFC stack temperature over the threshold.
The polynomial responses were then used to perform a Montecarlo simulation of the system, assuming that the system would work for one third of the total operating hours per year at each different load. The economic uncertainties related to fuel and electricity prices were considered at this level, assuming no particular correlation amongst them. This analysis allowed to evaluate the probability density function of the revenues of the plant, in particular the internal rate of return and the payback period. The resulting probability density function showed that, as expected, the variability of the prices of fuel and electricity has a strong impact on both IRR and PBP, leading to a coefficient of variance of about 10% for both parameters. The results suggested that, to mitigate the impact of technological uncertainties, allowances should be considered at the design phase mainly for the SOFC surface and recuperator surface, as well as a proper evaluation of the prices along the years. This would allow to reach the design efficiency and/or design power output at commissioning, alleviating at the same time the impact of fuel cell degradation along the projected life of SOFC stack.

- Finally, a methodology to include the uncertainties of market forecast within the **long-term thermo-economic assessment of power plants**. A test case was developed on a new solution for combined heat and power (CHP) generation, integrating a Combined Cycle Gas Turbine, CCGT, with a flue gas condensing heat pump (HP), and comparing it with a conventional coupling of the same

CCGT with a Heat-Only Boiler (HOB). To properly evaluate its feasibility from the economic point of view a thermo-economic analysis under uncertainty of the system was performed, modeling the cost of gas and single national price of electricity with a series of probability density functions to recreate the daily profile. A similar approach is followed for the gas cost, where the uncertainties related to the forecast capability are considered by creating a normal distribution with a proper standard deviation.
A Monte-Carlo simulation was performed on the two systems, and the outputs were then analyzed, leading to the following main conclusions:

- despite a higher annual cash flow for the CCGT-HP solution, the higher investment costs of this solution ended up in a lower NPV when compared to the CCGT-HOB solution. However, a lower variability of the CCGT-HP solution was also highlighted, ending in a more robust solution, especially in the more critical situation, the one with decreasing electrical prices.
- From the environmental point of view the CO2 emission was analyzed. As concerns this variable, the CCGT-HP presents lower emission than the CCGT-HOB, because the thermal capacity of the CHP power plant is extended exploiting the flue gas latent heat by the heat pump, thus increasing global efficiency, rather than burning additional fuel as in the HOB.

The present work illustrated how, especially in complex systems, model uncertainty can result in higher uncertainty in the outputs. Therefore, a stochastic analysis is generally important to evaluate the confidence on the results obtained from the simulations. Moreover, the uncertainties could also impact the system from an economic point of view, meaning that ignoring them could lead to fuzzy results. In spite of the challenges underlined in this work, probabilistic UQ is naturally amenable to mathematical and numerical analysis and, therefore, is expected to achieve a high level of maturity in the near future.

7.2 Future work

There may be some potential topics that can be conducted as future work in the area of uncertainty quantification for advanced energy systems. The methods analysed in this thesis provide a contribution to the treatment of uncertainties in the context of engineering design and proved their efficiency in particular in hybrid system analysis.
The last part of the dissertation (Section 6) has laid foundation in the field of uncertainty quantification applied to more complex energy systems. In this direction it could be interesting to perform an analysis which would also consider the uncertainties related to ambient conditions, such as ambient temperature and relative humidity, as they directly impact the performance of the system and cannot be determined precisely when performing a long-term analysis (as the expected life of such system is at least 20 years). The introduction of such uncertainties would require a proper statistical study on the historical data to be able to create a model for the creation of hourly temperature and humidity profiles.

Another important aspect would be the so-called "inverse analysis". This type of approach could be useful when, starting from the required probability distribution of the outputs (i.e. the maximum confidence band or variance allowed on the results, such as PBP of 3 years plus minus 6 months), we would like to know which is the maximum allowable standard deviation of some inputs, selected by the user.
Basing on the knowledge and experience gained from the present research, it results that uncertainty quantification techniques are actually improving the understanding of the advanced energy systems proposed in this work. In particular, the same techniques could be extended to broad areas in engineering such as model predictive control, complex innovative cycles which also need optimization routines.

Bibliography

[1] L. L. Green, H. Z. Lin, and M. R. Khalessi. "Probabilistic methods for uncertainty propagation applied to aircraft design". In: *20th AIAA Applied Aerodynamics Conference*. 2002. ISBN: 9781624101106. DOI: 10.2514/6.2002-3140.

[2] J. C. Helton and J. D. Johnson. "Quantification of margins and uncertainties: Alternative representations of epistemic uncertainty". In: *Reliability Engineering and System Safety* 96.9 (2011), pp. 1034–1052. ISSN: 09518320. DOI: 10.1016/j.ress.2011.02.013.

[3] H.-J. Zimmermann. *Fuzzy Set Theory—and Its Applications*. 2001. DOI: 10.1007/978-94-010-0646-0.

[4] H. R. Bae, R. V. Grandhi, and R. A. Canfield. "An approximation approach for uncertainty quantification using evidence theory". In: *Reliability Engineering and System Safety* (2004). ISSN: 09518320. DOI: 10.1016/j.ress.2004.01.011.

[5] H. R. Bae, R. V. Grandhi, and R. A. Canfield. "Sensitivity analysis of structural response uncertainty propagation using evidence theory". In: *Structural and Multidisciplinary Optimization* 31.4 (2006), pp. 270–279. ISSN: 1615147X. DOI: 10.1007/s00158-006-0606-9.

[6] S. F. Wojtkiewicz, M. S. Eldred, R. V. Field, A. Urbina, and J. R. Red-Horse. "Uncertainty quantification in large computational engineering models". In: *19th AIAA Applied Aerodynamics Conference*. 2001. DOI: 10.2514/6.2001-1455.

[7] C. J. Freitas. "The issue of numerical uncertainty". In: *Applied Mathematical Modelling* 26.2 (2002), pp. 237–248. ISSN: 0307904X. DOI: 10.1016/S0307-904X(01)00058-0.

[8] D. Draper. "Assessment and Propagation of Model Uncertainty". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1 (1995), pp. 45–70. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1995.tb02015.x.

[9] AIAA. "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations". In: *AIAA Journal* (1998). ISSN: 03064379. DOI: 10.2514/4.472855.001.

[10] A. Cullen and H. Frey. *Probabilistic Techniques in Exposure Assessment. A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*. 1999. DOI: 10.1002/sim.958.

[11] L. Todd. *The Oxford Handbook of Quantitative Methods, Volume 2 Statistical analysis*. 2013. ISBN: 9788578110796. arXiv: arXiv:1011.1669v3.

[12] P. R. Bevington, D. K. Robinson, J. M. Blair, A. J. Mallinckrodt, and S. McKay. "Data Reduction and Error Analysis for the Physical Sciences". In: *Computers in Physics* 7.4 (1993), p. 415. ISSN: 08941866. DOI: 10.1063/1.4823194.

[13] W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin. "Error and uncertainty in modeling and simulation". In: *Reliability Engineering and System Safety* 75.3 (2002), pp. 333–357. ISSN: 09518320. DOI: 10.1016/S0951-8320(01)00120-X.

[14] A. D. Kiureghian and O. Ditlevsen. "Aleatory or epistemic? Does it matter?" In: *Structural Safety* 31.2 (2009), pp. 105–112. ISSN: 01674730. DOI: 10.1016/j.strusafe.2008.06.020.

[15] O. De Weck, C. Eckert, and J. Clarkson. "A classification of uncertainty for early product and system design". In: *Proceedings of ICED 2007, the 16th International Conference on Engineering Design*. Vol. DS 42. 2007. ISBN: 1904670024.

[16] M. Phadke. *Quality Engineering Using Robust Design*. 1995, p. 250. DOI: 10.2307/1269049.

[17] R. L. Winkler. "Uncertainty in probabilistic risk assessment". In: *Reliability Engineering and System Safety* 54.2-3 (1996), pp. 127–132. ISSN: 09518320. DOI: 10.1016/S0951-8320(96)00070-1.

[18] R. Ghanem, H. Owhadi, and D. Higdon. *Handbook of uncertainty quantification*. Ed. by Springer. Berlin, Germany, 2017, pp. 1–2053. ISBN: 9783319123851. DOI: 10.1007/978-3-319-12385-1.

[19] Alain Desrosieres. *The Politics of Large Numbers: A History of Statistical Reasoning*. Cambridge, Massachussetts, and London, England, 1998. ISBN: 0674689321.

[20] B. J. Winer. *Statistical principles in experimental design*. 2009. DOI: 10.1037/11774-000.

[21] S. G. Kwak and J. H. Kim. "Central limit theorem: The cornerstone of modern statistics". In: *Korean Journal of Anesthesiology* 70.2 (2017), pp. 144–156. ISSN: 20057563. DOI: 10.4097/kjae.2017.70.2.144.

[22] G. J. Hahn and S. S. Shapiro. *Statistical Models in Engineering*. Wiley Classic Library, 1994.

[23] A. Cuneo, A. Traverso, and S. Shahpar. "Comparative analysis of methodologies for uncertainty propagation and quantification". In: *Proceedings of the ASME Turbo Expo*. Vol. 2C-2017. American Society of Mechanical Engineers, 2017. ISBN: 9780791850800. DOI: 10.1115/GT2017-63238. URL: https://asmedigitalcollection.asme.org/GT/proceedings/GT2017/50800/Charlotte,NorthCarolina,USA/241782.

[24] M. Morgan, M. H. Granger, and M. Small. *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Vol. 8. 1. 1990. ISBN: 0521427444. DOI: 10.1016/0169-2070(92)90021-z.

[25] J. C. Helton. "Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal". In: *Reliability Engineering and System Safety* 42.2-3 (1993), pp. 327–367. ISSN: 09518320. DOI: 10.1016/0951-8320(93)90097-I.

[26] I. M. Sobol. "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates". In: *Mathematics and Computers in Simulation* 55.1-3 (2001), pp. 271–280. ISSN: 03784754. DOI: 10.1016/S0378-4754(00)00270-6.

[27] N. Metropolis and S. Ulam. "The Monte Carlo Method". In: *Journal of the American Statistical Association* 44.247 (1949), pp. 335–341. ISSN: 1537274X. DOI: 10.1080/01621459.1949.10483310.

[28] N. Wiener. "The Homogeneous Chaos". In: *American Journal of Mathematics* 60.4 (1938), p. 897. ISSN: 00029327. DOI: 10.2307/2371268.

[29] H. Niederreiter. "Quasi-Monte Carlo methods and pseudo-random numbers". In: *Bulletin of the American Mathematical Society* 84.6 (1978), pp. 957–1042. ISSN: 0002-9904. DOI: 10.1090/s0002-9904-1978-14532-7.

[30] M. D. McKay, R. J. Beckman, and W. J. Conover. "Comparison of three methods for selecting values of input variables in the analysis of output from a computer code". In: *Technometrics* 21.2 (1979), pp. 239–245. ISSN: 15372723. DOI: 10.1080/00401706.1979.10489755.

[31] M. D. Dettinger and J. L. Wilson. "First order analysis of uncertainty in numerical models of groundwater flow part: 1. Mathematical development". In: *Water Resources Research* 17.1 (1981), pp. 149–161. ISSN: 19447973. DOI: 10.1029/WR017i001p00149.

[32] V. V. Toropov and V. L. Markine. "The use of simplified numerical models as mid-range approximations". In: *6th Symposium on Multidisciplinary Analysis and Optimization*. 1996, pp. 952–958. DOI: 10.2514/6.1996-4088.

[33] T. Gerstner and M. Griebel. "Numerical integration using sparse grids". In: *Numerical algorithms* 18.3 (1998), pp. 209–232. ISSN: 1017-1398. DOI: 10.1023/A:1019129717644.

[34] D. Xiu and G. Em Karniadakis. "Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos". In: *Computer Methods in Applied Mechanics and Engineering* 191.43 (2002), pp. 4927–4948. ISSN: 00457825. DOI: 10.1016/S0045-7825(02)00421-8.

[35] K. Subramanyan and U. M. Diwekar. "Characterization and quantification of uncertainty in solid oxide fuel cell hybrid power plants". In: *Journal of Power Sources* 142.1-2 (2005), pp. 103–116. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2004.09.030.

[36] U. M. Diwekar, E. S. Rubin, and H. C. Frey. "Optimal design of advanced power systems under uncertainty". In: *Energy Conversion and Management* 38.15-17 (1997), pp. 1725–1735. ISSN: 01968904. DOI: 10.1016/s0196-8904(96)00184-7.

[37] M. Q. Suo, Y. P. Li, and G. H. Huang. "Multicriteria decision making under uncertainty: An advanced ordered weighted averaging operator for planning electric power systems". In: *Engineering Applications of Artificial Intelligence* 25.1 (2012), pp. 72–81. ISSN: 09521976. DOI: 10.1016/j.engappai.2011.08.007.

[38] E. K. Hart and M. Z. Jacobson. "A Monte Carlo approach to generator portfolio planning and carbon emissions assessments of systems with large penetrations of variable renewables". In: *Renewable Energy* (2011). ISSN: 09601481. DOI: 10.1016/j.renene.2011.01.015.

[39] R. S. Gorla. "Probabilistic analysis of a solid-oxide fuel-cell based hybrid gas-turbine system". In: *Applied Energy* 78.1 (2004), pp. 63–74. ISSN: 03062619. DOI: 10.1016/S0306-2619(03)00006-0.

[40] A. Giugno, A. Cuneo, and A. Traverso. "Analysis of uncertainties in compact plate-fin recuperators for microturbines". In: *Applied Thermal Engineering* 150.September 2018 (2019), pp. 1243–1251. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2019.01.093. URL: https://doi.org/10.1016/j.applthermaleng.2019.01.093.

[41] A. Cuneo, L. Mantelli, A. Giugno, and A. Traverso. "Uncertainty Quantification Analysis of a Pressurised Fuel Cell Hybrid System". In: *Proceedings of ASME Turbo Expo 2019: Turbomachinery Technical Conference and Exposition*. Vol. 91351. Phoenix, 2019, pp. 1–15.

[42] A. Giugno, L. Mantelli, A. Cuneo, and A. Traverso. "Performance analysis of a fuel cell hybrid system subject to technological uncertainties". In: *Applied Energy* 279 (2020). ISSN: 03062619. DOI: 10.1016/j.apenergy.2020.115785.

[43] F. Montomoli, M. Massini, S. Salvadori, and F. Martelli. "Geometrical uncertainty and film cooling: Fillet radii". In: *Journal of Turbomachinery* 134.1 (2011). ISSN: 0889504X. DOI: 10.1115/1.4003287.

[44] A. Panizza, D. T. Rubino, and L. Tapinassi. "Efficient uncertainty quantification of centrifugal compressor performance using polynomial chaos". In: *Proceedings of the ASME Turbo Expo*. Vol. 2B. 2014. ISBN: 9780791845615. DOI: 10.1115/GT2014-25081.

[45] S. Abraham, M. Raisee, G. Ghorbaniasl, F. Contino, and C. Lacor. "A robust and efficient stepwise regression method for building sparse polynomial chaos expansions". In: *Journal of Computational Physics* 332 (2017), pp. 461–474. ISSN: 10902716. DOI: 10.1016/j.jcp.2016.12.015.

[46] S. Abraham, P. Tsirikoglou, C. Lacor, G. Ghorbanias, D. Wunsch, C. Hirsch, and F. Contino. "Uncertainty quantification in industrial turbo-machinery design using sparse polynomial chaos expansions". In: *2018 Multidisciplinary Analysis and Optimization Conference*. 2018. ISBN: 9781624105500. DOI: 10.2514/6.2018-3103.

[47] T. Ghisu, G. T. Parks, J. P. Jarrett, and P. J. Clarkson. "Adaptive polynomial chaos for gas turbine compression systems performance analysis". In: *AIAA Journal* (2010). ISSN: 00011452. DOI: 10.2514/1.J050012.

[48] D. Coppitters, W. De Paepe, and F. Contino. "Surrogate-assisted robust design optimization and global sensitivity analysis of a directly coupled photovoltaic-electrolyzer system under techno-economic uncertainty". In: *Applied Energy* 248 (2019), pp. 310–320. ISSN: 03062619. DOI: 10.1016/j.apenergy.2019.04.101.

[49] S. Shahpar and S. Caloni. "Aerodynamic optimisation of high pressure turbines for lean-burn combustion system". In: *Proceedings of the ASME Turbo Expo*. 2012. ISBN: 9780791844748. DOI: 10.1115/GT2012-69228.

[50] Y. M. Korolev, V. V. Toropov, and S. Shahpar. "Design optimization under uncertainty using the multipoint approximation method". In: *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2017*. 2017. ISBN: 9781624104534. DOI: 10.2514/6.2017-1934.

[51] A. K. Gupta, T. F. Móri, and G. J. Székely. "How to transform correlated random variables into uncorrelated ones". In: *Applied Mathematics Letters* 13.6 (2000), pp. 31–33. ISSN: 08939659. DOI: 10.1016/S0893-9659(00)00050-1.

[52] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf. "Global sensitivity analysis for systems with independent and/or correlated inputs". In: *Journal of Physical Chemistry A* 114.19 (2010), pp. 6022–6032. ISSN: 10895639. DOI: 10.1021/jp9096919.

[53] M. Rosenblatt. "Remarks on a Multivariate Transformation". In: *The Annals of Mathematical Statistics* 23.3 (1952), pp. 470–472. ISSN: 0003-4851. DOI: 10.1214/aoms/1177729394.

[54] H. Li and D. Zhang. "Probabilistic collocation method for flow in porous media: Comparisons with other stochastic methods". In: *Water Resources Research* 43.9 (2007). ISSN: 00431397. DOI: 10.1029/2006WR005673.

[55] J. L. Lumley. "The structure of inhomogeneous turbulence". In: *Atmospheric Turbulence and Radio Wave Propagation, edited by A. M. Yaglom and V. I. Tatarski (Nauka, Moscow)*. 1967, pp. 166–178. ISBN: 9783937655239.

[56] M. Eldred, C. Webster, and P. Constantine. "Evaluation of Non-Intrusive Approaches for Wiener-Askey Generalized Polynomial Chaos". In: 2008. DOI: 10.2514/6.2008-1892.

[57] K. Kim, M. R. Von Spakovsky, M. Wang, and D. J. Nelson. "Dynamic optimization under uncertainty of the synthesis/design and operation/control of a proton exchange membrane fuel cell system". In: *Journal of Power Sources* (2012). ISSN: 03787753. DOI: 10.1016/j.jpowsour.2011.11.014.

[58] D. Bigoni, A. P. Engsig-Karup, and H. True. "Comparison of classical and modern uncertainty qualification methods for the calculation of critical speeds in railway vehicle dynamics". In: *Proceedings of the Mini Conference on Vehicle System Dynamics, Identification and Anomalies*. Vol. 2012-Novem. 2012, pp. 91–100. ISBN: 9789633131022.

[59] K. P. Burnham and B. Efron. "The Jackknife, the Bootstrap and Other Resampling Plans." In: *Biometrics* 39.3 (1983), p. 816. ISSN: 0006341X. DOI: 10.2307/2531123.

[60] M. R. Chernick. *Bootstrap Methods*. 2007. DOI: 10.1002/9780470192573.

[61] C. Lemieux. *Monte Carlo and Quasi-Monte Carlo Sampling*. 2009. ISBN: 9780387781648. arXiv: arXiv:1011.1669v3.

[62] J. H. Halton. "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals". In: *Numerische Mathematik* 2.1 (1960), pp. 84–90. ISSN: 0029599X. DOI: 10.1007/BF01386213.

[63] J. M. Hammersley. "Monte Carlo Methods for Solving Multivariable Problems". In: *Annals of the New York Academy of Sciences* 86.3 (1960), pp. 844–874. ISSN: 17496632. DOI: 10.1111/j.1749-6632.1960.tb42846.x.

[64] J Peter and M Marcelet. "Comparison of surrogate models for turbomachinery design". In: *WSEAS Transactions on Fluid Mechanics* 3.1 (2008), pp. 10–17. URL: http://dl.acm.org/citation.cfm?id=1353862.1353870.

[65] P. L'Ecuyer. "Quasi-monte Carlo methods in finance". In: *Proceedings - Winter Simulation Conference*. Vol. 2. 2004, pp. 1645–1655. DOI: 10.1109/wsc.2004.1371512.

[66] C. Joy, P. P. Boyle, and K. S. Tan. "Quasi-Monte Carlo methods in numerical finance". In: *Management Science* 42.6 (1996), pp. 926–938. ISSN: 00251909. DOI: 10.1287/mnsc.42.6.926.

[67] B. J. Collings and H. Niederreiter. "Random Number Generation and Quasi-Monte Carlo Methods." In: *Journal of the American Statistical Association* 88.422 (1993), p. 699. ISSN: 01621459. DOI: 10.2307/2290359.

[68] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. 2004. ISBN: 978-0-387-76369-9.

[69] W. J. Morokoff and R. E. Caflisch. "Quasi-Monte carlo integration". In: *Journal of Computational Physics* 122.2 (1995), pp. 218–230. ISSN: 00219991. DOI: 10.1006/jcph.1995.1209.

[70] M. B. Giles. "Multilevel Monte Carlo path simulation". In: *Operations Research* 56.3 (2008), pp. 607–617. ISSN: 0030364X. DOI: 10.1287/opre.1070.0496.

[71] S. Mishra, C. Schwab, and J. Šukys. "Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions". In: *Journal of Computational Physics* 231.8 (2012), pp. 3365–3388. ISSN: 10902716. DOI: 10.1016/j.jcp.2012.01.011.

[72] D. C. Montgomery. *Design and Analysis of Experiments Eighth Edition*. 2012. ISBN: 9781118146927. DOI: 10.1198/tech.2006.s372.

[73] L. Cassettari, P. G. Giribone, M. Mosca, and R. Mosca. "The stochastic analysis of investments in industrial plants by simulation models with control of experimental error: Theory and application to a real business case". In: *Applied Mathematical Sciences* 4.73-76 (2010), pp. 3823–3840. ISSN: 1312885X.

[74] L. Cassettari, R. Mosca, and R. Revetria. "Monte Carlo simulation models evolving in replicated runs: A methodology to choose the optimal experimental sample size". In: *Mathematical Problems in Engineering* 2012 (2012), pp. 73–76. ISSN: 1024123X. DOI: 10.1155/2012/463873.

[75] R. F. Gunst, R. H. Myers, and D. C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. John Wiley & Sons, 2016. DOI: 10.2307/1270613.

[76] D. J. Pike, G. E. P. Box, and N. R. Draper. "Empirical Model-Building and Response Surfaces." In: *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 151.1 (1988), p. 223. ISSN: 09641998. DOI: 10.2307/2982196.

[77] L. Cassettari, R. Mosca, and R. Revetria. "Experimental error measurement in monte carlo simulation". In: *Handbook of Research on Discrete Event Simulation Environments: Technologies and Applications*. 2009, pp. 92–142. ISBN: 9781605667744. DOI: 10.4018/978-1-60566-774-4.ch006.

[78] U. Brussel. "Comparison of intrusive and non-intrusive polynomial chaos methods for CFD applications in aeronautics". In: *European Conference on Computational Fluid Dynamics* June (2010), pp. 14–17.

[79] R. G. Ghanem and P. D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. 1991. DOI: 10.1007/978-1-4612-3094-6.

[80] O. P. Le Maitre and O. M. Knio. *Spectral methods for uncertainty quantification*. 2010. ISBN: 9789048135196.

[81] D. Xiu and G. Em Karniadakis. "The Wiener-Askey polynomial chaos for stochastic differential equations". In: *SIAM Journal on Scientific Computing* 24.2 (2003), pp. 619–644. ISSN: 10648275. DOI: 10.1137/S1064827501387826.

[82] H. N. Najm. *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*. 2009. DOI: 10.1146/annurev.fluid.010908.165248.

[83] M. S. Eldred and J. Burkardt. "Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification". In: *47th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*. 2009. ISBN: 9781563479694. DOI: 10.2514/6.2009-976.

[84] I. Babuška, R. Temponet, and G. E. Zouraris. "Galerkin finite element approximations of stochastic elliptic partial differential equations". In: *SIAM Journal on Numerical Analysis* 42.2 (2004), pp. 800–825. ISSN: 00361429. DOI: 10.1137/S0036142902418680.

[85] S. Oladyshkin and W. Nowak. "Data-driven uncertainty quantification using the arbitrary polynomial chaos expansion". In: *Reliability Engineering and System Safety* 106 (2012), pp. 179–190. ISSN: 09518320. DOI: 10.1016/j.ress.2012.05.002.

[86] H. Scheffe. *The analysis of variance*. John Wiley & Sons, 1999.

[87] G. D. Ruxton. "The unequal variance t-test is an underused alternative to Student's t-test and the Mann-Whitney U test". In: *Behavioral Ecology* 17.4 (2006), pp. 688–690. ISSN: 10452249. DOI: 10.1093/beheco/ark016.

[88] T. K. Kim. "T test as a parametric statistic". In: *Korean Journal of Anesthesiology* 68.6 (2015), pp. 540–546. ISSN: 20057563. DOI: 10.4097/kjae.2015.68.6.540.

[89] S. P. Schacht, J. E. Aspelmeier, S. p. Schacht, and J. E. Aspelmeier. "One-Way Analysis of Variance (ANOVA)". In: *Social and Behavioral Statistics*. 2018, pp. 195–214. DOI: 10.4324/9780429497308-11.

[90] J. W. Tukey. "Comparing Individual Means in the Analysis of Variance". In: *Biometrics* 5.2 (1949), p. 99. ISSN: 0006341X. DOI: 10.2307/3001913.

[91] *F critical value for ANOVA test*. URL: https://web.ma.utexas.edu/users/davis/375/popecol/tables/f005.html (visited on 01/08/2021).

[92] T. K. Kim. "Understanding one-way anova using conceptual figures". In: *Korean Journal of Anesthesiology* 70.1 (2017), pp. 22–26. ISSN: 20057563. DOI: 10.4097/kjae.2017.70.1.22.

[93] M. Cavazzuti. *Optimization methods: From theory to design scientific and technological aspects in mechanics*. 2013, pp. 1–262. ISBN: 9783642311871. DOI: 10.1007/978-3-642-31187-1.

[94] C. C. Craig and R. A. Fisher. "The Design of Experiments." In: *The American Mathematical Monthly* 43.3 (1936), p. 180. ISSN: 00029890. DOI: 10.2307/2300364.

[95] M. I. Rodrigues and A. F. Iemma. *Experimental design and process optimization*. 2014. ISBN: 9781482299564. DOI: 10.1201/b17848.

[96] S. L. Ferreira et al. "Box-Behnken design: An alternative for the optimization of analytical methods". In: *Analytica Chimica Acta* 597.2 (2007), pp. 179–186. ISSN: 00032670. DOI: 10.1016/j.aca.2007.07.011.

[97] G. Tagushi, Y. Yokoyama, and Y. Wu. *Taguchi Methods: Design of Experiments*. 1993.

[98] H. Evangelaras, C. Koukouvinos, and M. V. Koutras. "Advances in Robust Parameter Design: From Taguchi's Inner-Outer Arrays to Combined Arrays". In: *Encyclopedia of Statistical Sciences*. 2011. DOI: 10.1002/0471667196.ess7146.

[99] W. L. Loh. "On latin hypercube sampling". In: *Annals of Statistics* 24.5 (1996), pp. 2058–2080. ISSN: 00905364. DOI: 10.1214/aos/1069362310.

[100] J. C. Helton and F. J. Davis. "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems". In: *Reliability Engineering and System Safety* 81.1 (2003), pp. 23–69. ISSN: 09518320. DOI: 10.1016/S0951-8320(03)00058-9.

[101] M. Stein. "Large sample properties of simulations using latin hypercube sampling". In: *Technometrics* 29.2 (1987), pp. 143–151. ISSN: 15372723. DOI: 10.1080/00401706.1987.10488205.

[102] A. Olsson, G. Sandberg, and O. Dahlblom. "On Latin hypercube sampling for structural reliability analysis". In: *Structural Safety* 25.1 (2003), pp. 47–68. ISSN: 01674730. DOI: 10.1016/S0167-4730(02)00039-5.

[103] L. N. Trefethen and D. Bau. *Numerical Linear Algebra*. 1997. DOI: 10.1137/1.9780898719574.

[104] G. W. S. and D. S. Watkins. "Fundamentals of Matrix Computations." In: *Mathematics of Computation* 59.199 (1992), p. 299. ISSN: 00255718. DOI: 10.2307/2153000.

[105] A. M. Law and W. D. Kelton. *Simulation modeling and analysis*. 1991. ISBN: 0780379241. DOI: 10.1016/j.sysconle.2007.02.002. arXiv: 0608329v1 [arXiv:astro-ph].

[106] N. Bradley. "Response Surface Methodology". PhD thesis. 2007. ISBN: 9780444527011. DOI: 10.1016/B978-044452701-1.00083-1.

[107] J. P. Kleijnen and R. G. Sargent. "A methodology for fitting and validating metamodels in simulation". In: *European Journal of Operational Research* 120.1 (2000), pp. 14–29. ISSN: 03772217. DOI: 10.1016/S0377-2217(98)00392-0.

[108] M. Mäkelä. "Experimental design and response surface methodology in energy applications: A tutorial review". In: *Energy Conversion and Management* 151 (2017), pp. 630–640. ISSN: 01968904. DOI: 10.1016/j.enconman.2017.09.021.

[109] J. P. Kleijnen. "An overview of the design and analysis of simulation experiments for sensitivity analysis". In: *European Journal of Operational Research* 164.2 (2005), pp. 287–300. ISSN: 03772217. DOI: 10.1016/j.ejor.2004.02.005.

[110] M. A. Bezerra, R. E. Santelli, E. P. Oliveira, L. S. Villar, and L. A. Escaleira. "Response surface methodology (RSM) as a tool for optimization in analytical chemistry". In: *Talanta* 76.5 (2008), pp. 965–977. ISSN: 00399140. DOI: 10.1016/j.talanta.2008.05.019.

[111] J. C. F. Gauss. *Combinationis observationum erroribus minimis obnoxiae*. University of Gottingen, 1825.

[112] M. A. Hariri-Ardebili, S. M. Seyed-Kolbadi, and M. Noori. "Response Surface Method for Material Uncertainty Quantification of Infrastructures". In: *Shock and Vibration* 2018 (2018). ISSN: 10709622. DOI: 10.1155/2018/1784203. URL: https://doi.org/10.1155/2018/1784203.

[113] E. H. Lockwood and A. L. Edwards. "An Introduction to Linear Regression and Correlation". In: *The Mathematical Gazette* 69.447 (1985), p. 62. ISSN: 00255572. DOI: 10.2307/3616472.

[114] N. Cressie. "The origins of kriging". In: *Mathematical Geology* 22.3 (1990), pp. 239–252. ISSN: 08828121. DOI: 10.1007/BF00889887.

[115] M. L. Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

[116] G. Bohling. *Introduction to geostatistics and variogram analysis*. October. 2005.

[117] N. Cressie. "Fitting variogram models by weighted least squares". In: *Journal of the International Association for Mathematical Geology* 17.5 (1985), pp. 563–586. ISSN: 00205958. DOI: 10.1007/BF01032109.

[118] Hiflux. *Hiflux recuperator*. URL: https://hiflux.co.uk/applications/microturbine-recuperators/.

[119] A. Traverso and A. F. Massardo. "Optimal design of compact recuperators for microturbine application". In: *Applied Thermal Engineering* 25.14-15 (2005), pp. 2054–2071. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2005.01.015.

[120] R. Shah. "Compact heat exchangers for microturbines". In: *Micro Gas Turbines* Educationa.September (2005), pp. 1–18. URL: http://services.bepress.com/cgi/viewcontent.cgi?article=1026{\&}context=eci/heatexchangerfall2005.

[121] G. Xiao, T. Yang, H. Liu, D. Ni, M. L. Ferrari, M. Li, Z. Luo, K. Cen, and M. Ni. *Recuperators for micro gas turbines: A review*. 2017. DOI: 10.1016/j.apenergy.2017.03.095.

[122] M. L. Ferrari, M. Pascenti, L. Magistri, and A. F. Massardo. "Micro gas turbine recuperator: Steady-state and transient experimental investigation". In: *Proceedings of the ASME Turbo Expo*. 2009. ISBN: 9780791848869. DOI: 10.1115/GT2009-59172.

[123] L. Galanti and A. F. Massardo. "Micro gas turbine thermodynamic and economic analysis up to 500 kWe size". In: *Applied Energy* 88.12 (2011), pp. 4795–4802. ISSN: 03062619. DOI: 10.1016/j.apenergy.2011.06.022.

[124] W. De Paepe, D. Coppitters, S. Abraham, P. Tsirikoglou, G. Ghorbaniasl, and F. Contino. "Robust Operational Optimization of a Typical micro Gas Turbine". In: *Energy Procedia*. Vol. 158. 2019, pp. 5795–5803. DOI: 10.1016/j.egypro.2019.01.549.

[125] C. F. McDonald. "Recuperator considerations for future higher efficiency microturbines". In: *Applied Thermal Engineering* 23.12 (2003), pp. 1463–1487. ISSN: 13594311. DOI: 10.1016/S1359-4311(03)00083-8.

[126] M. L. Ferrari, A. Sorce, M. Pascenti, and A. F. Massardo. "Recuperator dynamic performance: Experimental investigation with a microgas turbine test rig". In: *Applied Energy* 88.12 (2011), pp. 5090–5096. ISSN: 03062619. DOI: 10.1016/j.apenergy.2011.07.016.

[127] P. Maghsoudi, S. Sadeghi, H. Khanjarpanah, and H. H. Gorgani. "A comprehensive thermo-economic analysis, optimization and ranking of different microturbine plate-fin recuperators designs employing similar and dissimilar fins on hot and cold sides with NSGA-II algorithm and DEA model". In: *Applied Thermal Engineering* 130 (2018), pp. 1090–1104. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2017.11.087.

[128] B. J. Tsai and Y. L. Wang. "A novel Swiss-Roll recuperator for the microturbine engine". In: *Applied Thermal Engineering* 29.2-3 (2009), pp. 216–223. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2008.02.028.

[129] T. Stevens, F. Verplatesen, and M. Baelmans. "Requirements for recuperators in micro gas turbines". In: *PowerMEMS Conference*. 2004, pp. 96–99.

[130] G. Lagerström and M. Xie. "High performance & cost effective recuperator for micro-gas turbines". In: *American Society of Mechanical Engineers, International Gas Turbine Institute, Turbo Expo (Publication) IGTI*. 2002. DOI: 10.1115/gt2002-30402.

[131] E. Utriainen and B. Sundén. "A comparison of some heat transfer surfaces for small gas turbine recuperators". In: *Proceedings of the ASME Turbo Expo*. 2001. ISBN: 9780791878521. DOI: 10.1115/2001-GT-0474.

[132] E. Avallone, T. Baumeister, and A. Sadegh. *Marks' standard handbook for mechanical engineers*. 2007. DOI: 10.5860/CHOICE.44-6870. URL: http://choicereviews.org/review/10.5860/CHOICE.44-6870.

[133] W. M. Kays and A. L. London. *Compact heat exchangers. Third Edition*. 1984. ISBN: 0070334188.

[134] R. M. Manglik and A. E. Bergles. "Heat transfer and pressure drop correlations for the rectangular offset strip fin compact heat exchanger". In: *Experimental Thermal and Fluid Science* 10.2 (1995), pp. 171–180. ISSN: 08941777. DOI: 10.1016/0894-1777(94)00096-Q.

[135] C. F. McDonald. "Low-cost compact primary surface recuperator concept for microturbines". In: *Applied Thermal Engineering* 20.5 (2000), pp. 471–497. ISSN: 13594311. DOI: 10.1016/S1359-4311(99)00033-2.

[136] M. L. Ferrari, M. De Campo, and L. Magistri. "Design and emulation of a turbocharged bio-fuelled sofc plant". In: *Proceedings of the ASME Turbo Expo*. 2018. ISBN: 9780791851043. DOI: 10.1115/GT201875026.

[137] X. Zhang, S. H. Chan, G. Li, H. K. Ho, J. Li, and Z. Feng. "A review of integration strategies for solid oxide fuel cells". In: *Journal of Power Sources* 195.3 (2010), pp. 685–702. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2009.07.045.

[138] A. B. Stambouli and E. Traversa. "Solid oxide fuel cells (SOFCs): A review of an environmentally clean and efficient source of energy". In: *Renewable and Sustainable Energy Reviews* 6.5 (2002), pp. 433–455. ISSN: 13640321. DOI: 10.1016/S1364-0321(02)00014-X.

[139] V. Mehta and J. S. Cooper. "Review and analysis of PEM fuel cell design and manufacturing". In: *Journal of Power Sources* 114.1 (2003), pp. 32–53. ISSN: 03787753. DOI: 10.1016/S0378-7753(02)00542-6.

[140] M. A. Azizi and J. Brouwer. "Progress in solid oxide fuel cell-gas turbine hybrid power systems: System design and analysis, transient operation, controls and optimization". In: *Applied Energy* 215 (2018), pp. 237–289. ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.01.098.

[141] A. Cuneo, V. Zaccaria, D. Tucker, and A. Sorce. "Gas turbine size optimization in a hybrid system considering SOFC degradation". In: *Applied Energy* 230 (2018), pp. 855–864. ISSN: 03062619. DOI: 10.1016/j.apenergy.2018.09.027.

[142] M. Henke, J. Kallo, K. A. Friedrich, and W. G. Bessler. "Influence of pressurisation on SOFC performance and durability: A theoretical study". In: *Fuel Cells*. 2011. DOI: 10.1002/fuce.201000098.

[143] K. Venkataraman, E. C. Wanat, and L. D. Schmidt. "Steam reforming of methane and water-gas shift in catalytic wall reactors". In: *AIChE Journal* 49.5 (2003), pp. 1277–1284. ISSN: 00011541. DOI: 10.1002/aic.690490518.

[144] STAXERA. *SOFC Stack - Mk200, Product Data Sheet*. Tech. rep.

[145] L. Mantelli, M. L. Ferrari, and L. Magistri. "Off-design performance analysis of a turbocharged solid oxide fuel cell system". In: *Applied Thermal Engineering* 183.P1 (2021), pp. 116–134. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2020.116134. URL: https://doi.org/10.1016/j.applthermaleng.2020.116134.

[146] A. Traverso, A. F. Massardo, and R. Scarpellini. "Externally Fired micro-Gas Turbine: Modelling and experimental performance". In: *Applied Thermal Engineering* 26.16 (2006), pp. 1935–1941. ISSN: 13594311. DOI: 10.1016/j.applthermaleng.2006.01.013.

[147] M. L. Ferrari, M. Pascenti, and A. F. Massardo. "Ejector Model for High Temperature Fuel Cell Hybrid Systems: Experimental Validation at Steady-State and Dynamic Conditions". In: *Journal of Fuel Cell Science and Technology* 5.4 (2008), pp. 041005–1–7. ISSN: 1550624X. DOI: 10.1115/1.2890102.

[148] A. Greco, A. Sorce, R. Littwin, P. Costamagna, and L. Magistri. "Reformer faults in SOFC systems: Experimental and modeling analysis, and simulated fault maps". In: *International Journal of Hydrogen Energy* 39.36 (2014), pp. 21700–21713. ISSN: 03603199. DOI: 10.1016/j.ijhydene.2014.09.063.

[149] A. Sorce, A. Greco, L. Magistri, and P. Costamagna. "FDI oriented modeling of an experimental SOFC system, model validation and simulation of faulty states". In: *Applied Energy* 136 (2014), pp. 894–908. ISSN: 03062619. DOI: 10.1016/j.apenergy.2014.03.074.

[150] M. L. Ferrari. "Advanced control approach for hybrid systems based on solid oxide fuel cells". In: *Applied Energy* 145 (2015), pp. 364–373. ISSN: 03062619. DOI: 10.1016/j.apenergy.2015.02.059.

[151] R. L. Wasserstein and N. A. Lazar. *The ASA's Statement on p-Values: Context, Process, and Purpose*. 2016. DOI: 10.1080/00031305.2016.1154108.

[152] A. Cuneo, V. Zaccaria, D. Tucker, and A. Traverso. "Probabilistic analysis of a fuel cell degradation model for solid oxide fuel cell and gas turbine hybrid systems". In: *Energy* 141 (2017), pp. 2277–2287. ISSN: 03605442. DOI: 10.1016/j.energy.2017.12.002.

[153] O. Harel. "The estimation of R2 and adjusted R2 in incomplete data sets using multiple imputation". In: *Journal of Applied Statistics* 36.10 (2009), pp. 1109–1118. ISSN: 02664763. DOI: 10.1080/02664760802553000.

[154] A. Giugno, L. Mantelli, and A. Traverso. "Robust Design of a Fuel Cell-Turbocharger Hybrid Energy System". In: *Proceedings of the ASME Turbo Expo*. 2020.

[155] A. Arsalis. "Thermoeconomic modeling and parametric study of hybrid SOFC-gas turbine-steam turbine power plants ranging from 1.5 to 10 MWe". In: *Journal of Power Sources* 181.2 (2008), pp. 313–326. ISSN: 03787753. DOI: 10.1016/j.jpowsour.2007.11.104.

[156] A. Giugno, L. Mantelli, A. Cuneo, and A. Traverso. "Robust Design of a Hybrid Energy System". In: *E3S Web of Conferences*. Vol. 113. 2019. DOI: 10.1051/e3sconf/201911302008.

[157] Chemical Engineering. *CEPCI Archives - Chemical Engineering*. URL: https://www.chemengonline.com/tag/cepci/ (visited on 01/15/2021).

[158] Eurostat - Statistical Office of the European Communities. *Electricity prices components for non-household consumers - annual data*. Tech. rep. 2019. URL: https://ec.europa.eu/eurostat/data/database.

[159] EIA - U.S. Energy Information Administration. *Average Price of Electricity to Ultimate Customers by End-Use Sector*. Tech. rep. 2019. URL: https://www.eia.gov/electricity/monthly/epm{_}table{_}grapher.php?t=epmt{_}5{_}6{_}a.

[160] Eurostat - Statistical Office of the European Communities. *Gas prices components for non-household consumers - annual data*. Tech. rep. 2019. URL: https://ec.europa.eu/eurostat/data/database.

[161] EIA - U.S. Energy Information Administration. *Natural Gas Prices*. Tech. rep. 2019. URL: https://www.eia.gov/dnav/ng/ng{_}pri{_}sum{_}dcu{_}SNY{_}a.htm.

[162] Statease. *Interpretation of R-squared*. URL: https://www.statease.com/docs/v11/contents/analysis/interpretation-of-r-squared/ (visited on 01/15/2021).

[163] T. Murata and H. Ishibuchi. "MOGA: multi-objective genetic algorithms". In: *Proceedings of the IEEE Conference on Evolutionary Computation*. Vol. 1. 1995, pp. 289–294. DOI: 10.1109/icec.1995.489161.

[164] C. M. Fonseca and P. J. Fleming. "Genetic Algorithms for Multiobjective Optimization :" in: *Icga* July (1993). ISSN: 14639076. DOI: citeulike-article-id:2361311.

[165] European Commission. *A policy framework for climate and energy in the period from 2020 to 2030*. Tech. rep. 2014.

[166] Eurostat. *Energy, transport and environment indicators. 2018 edition*. 2018, p. 232. ISBN: 978-92-79-96509-8. DOI: 10.2785/94549. URL: htttp://dx.doi.org/10.2785/94549.

[167] M. Minutillo, A. Perna, and A. Sorce. "Combined hydrogen, heat and electricity generation via biogas reforming: Energy and economic assessments". In: *International Journal of Hydrogen Energy* 44.43 (2019), pp. 23880–23898. ISSN: 03603199. DOI: 10.1016/j.ijhydene.2019.07.105.

[168] D. Bellotti, A. Sorce, M. Rivarolo, and L. Magistri. "Techno-economic analysis for the integration of a power to fuel system with a CCS coal power plant". In: *Journal of CO2 Utilization* 33 (2019), pp. 262–272. ISSN: 22129820. DOI: 10.1016/j.jcou.2019.05.019.

[169] D. Bellotti, L. Cassettari, M. Mosca, and L. Magistri. "RSM approach for stochastic sensitivity analysis of the economic sustainability of a methanol production plant using renewable energy sources". In: *Journal of Cleaner Production* 240 (2019). ISSN: 09596526. DOI: 10.1016/j.jclepro.2019.117947.

[170] A. D. Smith, P. J. Mago, and N. Fumo. "Benefits of thermal energy storage option combined with CHP system for different commercial building types". In: *Sustainable Energy Technologies and Assessments* 1.1 (2013), pp. 3–12. ISSN: 22131388. DOI: 10.1016/j.seta.2012.11.001.

[171] B. Caldecott and J. McDaniels. "Stranded generation assets : Implications for European capacity mechanisms , energy markets and climate policy". In: *Stranded Assets Programme, SSEE, University of Oxford* January (2014), pp. 1–62.

[172] S. Cornot-Gandolphe. *Gas and Coal Competition in the EU Power Sector*. 2014.

[173] R. Wiltshire. *Advanced District Heating and Cooling (DHC) Systems*. 2015, pp. 1–364. ISBN: 9781782423959. DOI: 10.1016/c2014-0-01422-0.

[174] A. David, B. V. Mathiesen, H. Averfalk, S. Werner, and H. Lund. "Heat Roadmap Europe: Large-scale electric heat pumps in district heating systems". In: *Energies* 10.4 (2017). ISSN: 19961073. DOI: 10.3390/en10040578.

[175] R. Lund and U. Persson. "Mapping of potential heat sources for heat pumps for district heating in Denmark". In: *Energy* 110 (2016), pp. 129–138. ISSN: 03605442. DOI: 10.1016/j.energy.2015.12.127.

[176] F. Levihn. "CHP and heat pumps to balance renewable power production: Lessons from the district heating network in Stockholm". In: *Energy* 137 (2017), pp. 670–678. ISSN: 03605442. DOI: 10.1016/j.energy.2017.01.118.

[177] M. B. Blarke and E. Dotzauer. "Intermittency-friendly and high-efficiency cogeneration: Operational optimisation of cogeneration with compression heat pump, flue gas heat recovery, and intermediate cold storage". In: *Energy* 36.12 (2011), pp. 6867–6878. ISSN: 03605442. DOI: 10.1016/j.energy.2011.10.008.

[178] T. Ommen, W. B. Markussen, and B. Elmegaard. "Heat pumps in combined heat and power systems". In: *Energy* 76 (2014), pp. 989–1000. ISSN: 03605442. DOI: 10.1016/j.energy.2014.09.016.

[179] Q. Chen, K. Finney, H. Li, X. Zhang, J. Zhou, V. Sharifi, and J. Swithenbank. "Condensing boiler applications in the process industry". In: *Applied Energy* 89.1 (2012), pp. 30–36. ISSN: 03062619. DOI: 10.1016/j.apenergy.2010.11.020.

[180] B. Hebenstreit, R. Schnetzinger, R. Ohnmacht, E. Höftberger, J. Lundgren, W. Haslinger, and A. Toffolo. "Techno-economic study of a heat pump enhanced flue gas heat recovery for biomass boilers". In: *Biomass and Bioenergy* 71 (2014), pp. 12–22. ISSN: 09619534. DOI: 10.1016/j.biombioe.2014.01.048.

[181] M. Qu, O. Abdelaziz, and H. Yin. "New configurations of a heat recovery absorption heat pump integrated with a natural gas boiler for boiler efficiency improvement". In: *Energy Conversion and Management* 87 (2014), pp. 175–184. ISSN: 01968904. DOI: 10.1016/j.enconman.2014.06.083.

[182] Danish Energy Agency. *Technology Data for Energy Plants for Electricity and District heating generation*. Tech. rep. June. 2019, pp. 1–374.

[183] A. Molyneaux, G. Leyland, and D. Favrat. "Environomic multi-objective optimisation of a district heating network considering centralized and decentralized heat pumps". In: *Energy* 35.2 (2010), pp. 751–758. ISSN: 03605442. DOI: 10.1016/j.energy.2009.09.028.

[184] *Pump-Heat*. URL: www.pumpheat.eu (visited on 01/27/2021).

[185] H. Lund, S. Werner, R. Wiltshire, S. Svendsen, J. E. Thorsen, F. Hvelplund, and B. V. Mathiesen. *4th Generation District Heating (4GDH). Integrating smart thermal grids into future sustainable energy systems.* 2014. DOI: 10.1016/j.energy.2014.02.089.

[186] A. Sorce, A. Giugno, D. Marino, S. Piola, and R. Guedez. "Analysis of a combined cycle exploiting inlet conditioning technologies for power modulation". In: *Proceedings of the ASME Turbo Expo*. Vol. 3. 2019, pp. 1–14. ISBN: 9780791858608. DOI: 10.1115/GT2019-91541.

[187] Friotherm. *Uniturbo 50FY – Centrifugal Compressor for Large Scale Refrigeration Plants and Heat Pumps*. Tech. rep.

[188] I. H. Bell, J. Wronski, S. Quoilin, and V. Lemort. "Pure and pseudo-pure fluid thermophysical property evaluation and the open-source thermophysical property library coolprop". In: *Industrial and Engineering Chemistry Research* 53.6 (2014), pp. 2498–2508. ISSN: 08885885. DOI: 10.1021/ie4033999.

[189] A. society of heating refrigerating and air conditioning engineers. *1997 ASHRAE Handbook: Fundamentals*. 1997, p. 851.

[190] H. Satyavada and S. Baldi. "A novel modelling approach for condensing boilers based on hybrid dynamical systems". In: *Machines* 4.2 (2016). ISSN: 20751702. DOI: 10.3390/machines4020010.

[191] M. S. Peters, K. D. Timmerhaus, and R. E. West. *Plant Design and Economics for Chemical Engineers 5th Edition*. 2003. ISBN: 9789896540821. arXiv: arXiv:1011.1669v3.

[192] P. Capros et al. *EU Reference Scenario 2016 - A Policy Framework for Climate and Energy in the Period from 2020 to 2030*. Tech. rep. 2016, p. 27. DOI: 10.2833/9127.

[193] Gestore dei Servizi Energetici. *Assessment of the National and Regional Potential for the Application of High-Efficiency Cogeneration and Efficient District Heating*. Tech. rep. 2016.

[194] Autorità Garante della Concorrenza e del Mercato. *Indagine Conoscitiva Sul Settore Del Teleriscaldamento*. Tech. rep. 2014.

[195] H. Li, J. Song, Q. Sun, F. Wallin, and Q. Zhang. "A dynamic price model based on levelized cost for district heating". In: *Energy, Ecology and Environment* 4.1 (2019), pp. 15–25. ISSN: 23638338. DOI: 10.1007/s40974-019-00109-6.

[196] Gestore Mercati Energetici. *Dati Storici Mercati*. Tech. rep. 2018.

[197] M. dell'ambiente e della tutela del territorio e del mare. *Decreto Legislativo n.30 Del 2013, Tabella Allegata Dei Parametri Standard Nazionali per Il Monitoraggio e La Comunicazione Dei Gas Ad Effetto Serra*. 2019.

[198] TERNA. *Dati statistici sull'energia elettrica in Italia*. Tech. rep. 2018. URL: http://www.terna.it/SistemaElettrico/StatisticheePrevisioni/DatiStatistici.aspx.

[199] R. A. Fisher. "On the mathematical foundations of theoretical statistics". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 222.594-604 (1922), pp. 309–368. ISSN: 0264-3952. DOI: 10.1098/rsta.1922.0009.

[200] European Commission. *EU Reference Scenario 2016 - Energy, transport and GHG emissions - Trends to 2050*. Tech. rep. 2016. URL: https://ec.europa.eu/energy/sites/ener/files/documents/20160713draft{_}publication{_}REF2016{_}v13.pdf.